



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>C12N 15/12, C07R 14/47, C12Q 1/68, A61R 38/17, C07K 16/18</b>		A2	(11) International Publication Number: <b>WO 00/23589</b>
			(43) International Publication Date: 27 April 2000 (27.04.00)
(21) International Application Number: PCT/US99/24511		(72) Inventors; and	
(22) International Filing Date: 19 October 1999 (19.10.99)		(75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #12, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue, #30, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). YANG, Junming [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US). SHIH, Leo, L. [US/US]; Apartment B, 1081 Tanland Drive, Palo Alto, CA 94304 (US).	
(30) Priority Data:			
60/172,216	20 October 1998 (20.10.98)	US	
60/118,559	4 February 1999 (04.02.99)	US	
60/172,229	11 February 1999 (11.02.99)	US	
60/154,336	22 April 1999 (22.04.99)	US	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications			
US	60/172,216 (CIP)		
Filed on	20 October 1998 (20.10.98)		
US	60/118,559 (CIP)		
Filed on	4 February 1999 (04.02.99)		
US	60/172,229 (CIP)		
Filed on	11 February 1999 (11.02.99)		
US	60/154,336 (CIP)		
Filed on	22 April 1999 (22.04.99)		
(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).		(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).	
		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
		Published Without international search report and to be republished upon receipt of that report.	
(54) Title: PROLIFERATION AND APOPTOSIS RELATED PROTEINS			
<p>1 MSRTMARTRPGOLG--RVTGAGGWGSAAVC 1342011 1 MATPVPPPSPRHLRLRLLLSG-----LI GI452276</p> <p>29 RGRALRGREPALPSASFDPVAACPGSLDCA 1342011 25 LGALALNG-----ATAARRPDATTCPGSLDCA GI452276</p> <p>59 LKRRARCPPGAHAACGPGCLQPFQEDQOGLCV 1342011 50 LKRRARKCPPGAHAACGPGCLSFOEDQRGFCV GI452276</p> <p>89 PRMRRPFGGGRPPQPRLEDEIDFLAQELA-- 1342011 80 PRKHLSSGEGLPQPRLEFEIDS LAQELA LK GI452276</p> <p>117 RKE SGHS--TPFLPKDRQRLPEPA-TLGF 1342011 110 EKELAGHSRLTAQPLLERAO KLLEPAATLGF GI452276</p> <p>143 SARGQGLELGLPSTPGTPTPTPHTSLQSPV 1342011 140 SQWGORLEFGLPSTHGTS SPILPHTSLSSRA GI452276</p> <p>173 SSDPVHMSPLEPRGGQGDGLALVLLILAF CV 1342011 170 SSGPVPQMSPLEPQGRHGNGLTLVLLILAF CL GI452276</p>			
(57) Abstract			
<p>The invention provides human proliferation and apoptosis related proteins (PROAP) and polynucleotides which identify and encode PROAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of PROAP.</p>			

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## PROLIFERATION AND APOPTOSIS RELATED PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proliferation and apoptosis related proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, immunological, and reproductive disorders.

5

### BACKGROUND OF THE INVENTION

Tissue growth involves complex and ordered patterns of cell proliferation, cell differentiation, and regulated cell death (apoptosis). Cell proliferation and apoptosis are regulated to maintain both the number and the spatial organization of cells. This regulation depends on appropriate expression of proteins which control cell cycle progression in response to extracellular signals, such as growth factors and other mitogens, and intracellular cues, such as DNA damage or nutrient starvation. Molecules which directly or indirectly modulate cell cycle progression fall into several categories, including growth factors and their receptors, second messenger and signal transduction proteins, oncogene products, tumor-suppressor proteins, and mitosis-promoting factors. Cancers are characterized by continuous or uncontrolled cell proliferation. Some cancers are associated with suppression of normal apoptotic cell death.

#### Growth Factors and Signal Transduction Machinery

Growth factors are typically large, secreted polypeptides that act on cells in their local environment to promote cell proliferation. Growth factors bind to and activate specific cell surface receptors that initiate intracellular signal transduction cascades. Many growth factor receptors are classified as receptor tyrosine kinases that undergo autophosphorylation upon ligand binding. Autophosphorylation enables the receptor to interact with signal transduction proteins such as SH2 or SH3 (Src homology regions 2 or 3) domain-containing proteins. Other proteins that act downstream of growth factor receptors contain unique signaling domains such as the SPRY (Spl a and ryanodine receptor) domain. (See, for example, Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864.) These proteins then modulate the activity state of small G-proteins, such as Ras, Rab, and Rho, along with GTPase activating proteins (GAPs), guanine nucleotide releasing proteins (GNRPs), and other guanine nucleotide exchange factors. Small G proteins act as molecular switches that turn on mitogen-activated protein kinase (MAP kinase) cascades. MAP kinase activates transcription of the early-response genes discussed below.

Most growth factors also have a multitude of other actions besides the regulation of cell growth and division: they can control the proliferation, survival, differentiation, migration, or function

of cells depending on the circumstance. For example, epidermal growth factor (EGF) protects gastric mucosa against injury and accelerates ulcer healing by stimulating cell migration and proliferation. EGF binds the transmembrane protein tyrosine kinase EGF-R to trigger a series of events that results in activation of the Ras/Raf/MAP kinase pathway by the GTP-binding protein Ras. Other pathways potentially activated by EGF include the phosphatidylinositol pathway and the JAK/STAT signaling pathway (Tarnawski, A.S. et al. (1998) *J. Clin. Gastroenterol.* 27:S12-S20).

In addition to growth factors, small signaling peptides and hormones also influence cell proliferation. These molecules bind primarily to another class of receptor, the trimeric G-protein coupled receptor (GPCR), found predominantly on the surface of immune, neuronal, and neuroendocrine cells. Upon ligand binding, the GPCR activates a trimeric G protein which in turn triggers increased levels of intracellular second messengers such as phospholipase C,  $Ca^{2+}$ , and cyclic AMP. Most GPCR-mediated signaling pathways indirectly promote cell proliferation by causing the secretion or breakdown of other signaling molecules that have direct mitogenic effects (Smith, A. et al. (1994) *Cell* 76:959-962).

Protein kinase C (PKC) plays a central role in the control of proliferation and differentiation of various cell types by mediating the signal transduction response to hormones and growth factors. The PKC family of serine/threonine kinases includes twelve different isoforms which have similar catalytic domains at their C-termini, but differ in their N-terminal regulatory domains. Since most cells express multiple PKC isoforms, the specificity of each enzyme for its substrate is achieved by targeting individual isoenzymes to a select location in the cell, either constitutively or upon cell stimulation. A variety of PKC-binding proteins and lipids have been identified that may function to compartmentalize PKC isoenzymes, including RACK1, serum deprivation response (sdr) protein, and SRBC (sdr-related gene product that binds C-kinase). Interestingly, both sdr and SRBC appear to provide localization of activated PKC to caveolae, but each has specificity for a different isoenzyme; sdr interacts specifically with PKC $\alpha$  and SRBC interacts with PKC $\delta$ . Both sdr and SRBC are induced during stages of growth arrest, and were originally isolated from serum-deprived cultured cells. Thus, sdr and SRBC appear to be important for targeting activated PKC isoenzymes to subcellular signaling sites important in growth control. (Mineo, C. et al. (1998) *J. Cell Biol.* 141:601-610; and Izumi, Y. et al. (1997) *J. Biol. Chem.* 272:7381-7389.)

### Oncogenes

Oncogenes (i.e. "cancer-causing genes") are involved in the reception and transduction of growth factor signals and in the modulation of gene expression in response to these signals. For example, stimulation of a cell by growth factor activates two sets of genes, the early-response genes and the delayed-response genes. Early-response gene products include myc, fos, and jun, all of



which encode gene regulatory proteins. These regulatory proteins activate the transcription of the delayed-response genes which encode proteins directly involved in cell cycle progression, such as the cyclins and cyclin dependent kinase discussed below. Additional oncogene products which directly regulate gene expression include the Rel transcription factor, the Ret zinc finger protein, and the Tre oncoprotein. (See, for example, Cao, T. et al. (1998) J. Cell Sci. 111:1319-1329; and Nakamura, T. et al. (1992) Oncogene 7:733-741.) Some conserved regions of oncogenes have been identified, such as the C3HC4 RING finger motif. Mutations in the C3HC4 RING finger domain of the Bmi-1 oncoprotein, for example, block lymphoma induction in mice (Hemenway, C.S. (1998) Oncogene 16:2541-2547). Apoptosis inhibition motifs have also been identified, such as the BIR repeat implicated in the activity of the IAP (Inhibitor of Apoptosis) family. Mutations or chromosomal translocations which result in hyperactivation of oncogenes result in uncontrolled cell proliferation.

#### Tumor Suppressors

Tumor suppressor genes are involved in inhibition of cell proliferation. Mutations which decrease the activity of tumor suppressor genes result in increased cell proliferation. In humans and other mammals, tumor suppressors include the retinoblastoma (Rb) and p53 proteins. Tumor suppressors have also been discovered in lower animals such as Drosophila, in which the Discs-Large (Dlg) and Hyperplastic Discs (Hyd) proteins inhibit hyperplasia of undifferentiated epithelial cells in developing imaginal discs. (See, for example, Mansfield, E. et al. (1994) Dev. Biol. 165:507-526.) The importance of tumor suppressor genes and oncogenes in the development of cancer is demonstrated by the fact that about 75% of colorectal cancers have inactivating mutations in the p53 gene and about 50% have a hyper-activating mutation in a ras family oncogene.

Tumor suppressor genes often act as "gatekeepers" (Kinzler, K.W. and Vogelstein, B. (1996) Cell 87:159-170). Normally, the gatekeeper is responsible for maintaining a balance of cell division, growth arrest, and death. External signals may activate or inactivate the gatekeeper, or alter its location within the cell. In some cases, inactivation of the gatekeeper is necessary for cell proliferation, and activation is necessary for cell growth arrest and differentiation. In other cases, the situation is reversed. Proteins which interact with the gatekeeper modify its activity or intracellular location to provide the appropriate response to external signals at any stage in the cell's development.

An example of a gatekeeper protein is the adenomatous polyposis coli (APC) protein. Though APC is expressed ubiquitously, it appears to function as a gatekeeper in colorectal cells. Mutations in the APC protein are linked to familial and sporadic forms of colon cancer. All of these mutations involve truncations in the APC C-terminus, which serves as a binding site for several proteins, including EB1, RPl, and the tumor suppressor protein Dlg. The interactions between APC and these binding proteins may be important for localizing or regulating APC activity. For example,

EB1 appears to link APC to microtubules, and a defect in chromosome segregation has been implicated as an early event in colorectal tumorigenesis (Berreuta, L. (1998) Proc. Natl. Acad. Sci. USA 95:10596-10601; and Renner, C. et al. (1997) J. Immunol. 159:1276-1283).

Another example of a gatekeeper is the E2F transcription factor, which can function either as a positive regulator of cell cycle progression or as a suppressor of cell proliferation, depending on the tissue. The balance of cell division over growth arrest and differentiation appears to involve proteins which interact with and modulate E2F. These proteins include the Rb tumor suppressor protein and NPDC-1 (neural proliferation, differentiation, and control). Rb acts to repress transcriptional activity of E2F, leading to differentiation or apoptosis in the responding cell. NPDC-1 is a neural specific gene expressed in growth arrested and differentiated cells. The NPDC-1 gene product, npdcf-1, interacts with E2F to down-regulate cell proliferation (Dupont, E. et al. (1998) J. Neurosci. Res. 51:257-267).

#### Cell Cycle Machinery

The molecular machinery which drives the cell cycle in response to mitogens and growth factors has been extensively studied in model systems such as budding yeast, fission yeast, and the African clawed frog, Xenopus. Essentially, the cell cycle is comprised of four successive phases: G1, S (DNA synthesis), G2, and M (mitosis). Cells which exit the cell cycle enter a quiescent phase called G0. Studies in yeast have shown that exit from S and M phases is driven by the anaphase-promoting complex, an assembly of proteins that degrades cyclins via the ubiquitin-mediated protein degradation pathway. (See, for example, Kominami, K. et al. (1998) EMBO J. 17:5388-5399.) Other non-kinase proteins, such as the Zer1p RNA splicing protein in fission yeast, are important for exit of the cell from G0 and entry into G1 or G2. (See, for example, Urushiyama, S. et al. (1997) Genetics 147:101-115.)

Several cell cycle transitions, including the entry and exit of a cell from mitosis, are dependent upon the activation and inhibition of cyclin-dependent kinases (Cdks). The Cdks are composed of a kinase subunit, Cdk, and an activating subunit, cyclin, in a complex that is subject to many levels of regulation. Cyclins bind and activate cyclin-dependent protein kinases which then phosphorylate and activate selected proteins involved in the mitotic process. The Cdk-cyclin complex is both activated and inhibited by phosphorylation. In addition, the Cdk-cyclin complex is regulated by targeted degradation involving molecules such as CDC4 and CDC53. Other proteins mediate entry into or progression through mitosis. For example, Berry and Gould recently identified a novel, 142 amino acid protein from the yeast S. pombe, termed dmp1p, that is required for proper spindle formation and entry into mitosis, but does not interact with cyclin-type proteins (Berry L.D. and Gould K.L. (1997) J. Cell Biol. 137:1337-1354). Dim1p appears to be evolutionarily conserved,

since a human homolog has recently been described (Larin D., et al. (1997) GI 2565275).

#### Apoptosis Machinery

Apoptosis is the genetically controlled process by which unneeded or defective cells undergo programmed cell death. Selective elimination of cells is as important for morphogenesis and tissue remodeling as is cell proliferation and differentiation. Lack of apoptosis may result in hyperplasia and other disorders associated with increased cell proliferation. Apoptosis is also a critical component of the immune response. Immune cells such as cytotoxic T-cells and natural killer cells prevent the spread of disease by inducing apoptosis in tumor cells and virus-infected cells. In addition, immune cells that fail to distinguish self molecules from foreign molecules must be eliminated by apoptosis to avoid an autoimmune response.

Apoptotic cells undergo distinct morphological changes. Hallmarks of apoptosis include cell shrinkage, nuclear and cytoplasmic condensation, and alterations in plasma membrane topology. Biochemically, apoptotic cells are characterized by increased intracellular calcium concentration, fragmentation of chromosomal DNA, and expression of novel cell surface components.

The molecular mechanisms of apoptosis are highly conserved, and many of the key protein regulators and effectors of apoptosis have been identified. Apoptosis generally proceeds in response to a signal which is transduced intracellularly and results in altered patterns of gene expression and protein activity. Signaling molecules such as hormones and cytokines are known both to stimulate and to inhibit apoptosis through interactions with cell surface receptors. Transcription factors also play an important role in the onset of apoptosis. A number of downstream effector molecules, particularly proteases such as the cysteine proteases called caspases, have been implicated in the degradation of cellular components and the proteolytic activation of other apoptotic effectors.

The Fas/Apo-1 receptor (FAS) is a member of the tumor necrosis factor-receptor family. Upon binding its ligand (Fas ligand), the membrane-spanning FAS induces apoptosis by recruiting several cytoplasmic proteins that transmit the death signal. Chu et al. isolated one such protein from mice, termed FAS-associated protein factor 1 (FAF1), and demonstrated that expression of FAF1 in L cells potentiated FAS-induced apoptosis (Chu, K. et al. (1995) Proc. Natl. Acad. Sci. USA 92:11894-11898). Subsequently, FAS-associated factors have been isolated from numerous other species, including quail and fly (Frohlich, T., et al. (1998) J. Cell Sci. 111:2353-63; and Lukacsovich, T. et al. (1998) GI 3688609).

Fragmentation of chromosomal DNA is one of the hallmarks of apoptosis. DNA fragmentation factor (DFF) is a protein composed of two subunits, a 40-kDa, caspase-activated nuclease termed DFF40/CAD, and its 45-kDa inhibitor DFF45/ICAD. Two mouse homologs of DFF45/ICAD, termed CIDE-A and CIDE-B, have recently been described (Inohara, N. et al. (1998)

EMBO J. 17:2526-2533). CIDE-A and CIDE-B expression in mammalian cells activated apoptosis, while expression of CIDE-A alone induced DNA fragmentation. In addition, FAS-mediated apoptosis was enhanced by CIDE-A and CIDE-B, further implicating these proteins as effectors that mediate apoptosis.

5           Cancers are characterized by inappropriate cell proliferation, which may be due to uncontrolled cell growth or inadequate apoptosis. Strategies for treatment may involve either reestablishing control over cell cycle progression, or selectively stimulating apoptosis in cancerous cells (Nigg, E.A. (1995) BioEssays 17:471-480).

          Immunological defenses against cancer include induction of apoptosis in mutant cells by  
10   tumor suppressors, and the recognition of tumor antigens by T lymphocytes. Response to mitogenic stresses is frequently controlled at the level of transcription and is coordinated by various transcription factors. The Rel/NF-kappa B family of vertebrate transcription factors, for example, plays a pivotal role in inflammatory and immune responses to radiation. The NF-kappa B family includes p50, p52, RelA, RelB, and cRel and other DNA-binding proteins. The p52 protein induces  
15   apoptosis, upregulates transcription factor c-Jun, and activates c-Jun N-terminal kinase 1 (JNK1) (Sun, L. et al. (1998) Gene 208:157-166). Most NF-kappa B proteins form DNA-binding homodimers or heterodimers. Dimerization of many transcription factors is mediated by a conserved sequence known as the bZIP domain, characterised by a basic region followed by a leucine zipper.

          The discovery of new proliferation and apoptosis related proteins and the polynucleotides  
20   encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, immunological, and reproductive disorders.

### SUMMARY OF THE INVENTION

          The invention features substantially purified polypeptides, proliferation and apoptosis related  
25   proteins, referred to collectively as "PROAP" and individually as "PROAP-1," "PROAP-2," "PROAP-3," "PROAP-4," "PROAP-5," "PROAP-6," "PROAP-7," "PROAP-8," "PROAP-9," "PROAP-10," "PROAP-11," "PROAP-12," "PROAP-13," "PROAP-14," "PROAP-15," "PROAP-16," "PROAP-17," "PROAP-18," and "PROAP-19." In one aspect, the invention provides a substantially purified polypeptide comprising an amino acid sequence selected from the group  
30   consisting of SEQ ID NO:1-19 and fragments thereof. The invention also includes a polypeptide comprising an amino acid sequence that differs by one or more conservative amino acid substitutions from an amino acid sequence selected from the group consisting of SEQ ID NO:1-19.

          The invention further provides a substantially purified variant having at least 90% amino acid identity to at least one of the amino acid sequences selected from the group consisting of SEQ ID

NO:1-19 and fragments thereof. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 and fragments thereof. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of  
5 SEQ ID NO:1-19 and fragments thereof.

Additionally, the invention provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19 and fragments thereof. The  
10 invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide encoding the polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:1-19 and fragments thereof.

The invention also provides a method for detecting a polynucleotide in a sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the  
15 polynucleotide sequence to at least one of the polynucleotides of the sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide in the sample. In one aspect, the method further comprises amplifying the polynucleotide prior to hybridization.

The invention also provides an isolated and purified polynucleotide comprising a  
20 polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38 and fragments thereof. The invention further provides an isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38 and fragments thereof. The invention also provides an isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide comprising  
25 a polynucleotide sequence selected from the group consisting of SEQ ID NO:20-38 and fragments thereof.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-19. In another aspect, the expression vector is contained within a host  
30 cell.

The invention also provides a method for producing a polypeptide, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing a polynucleotide of the invention under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-19 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

5 The invention further includes a purified antibody which binds to a polypeptide selected from the group consisting of SEQ ID NO:1-19 and fragments thereof. The invention also provides a purified agonist and a purified antagonist to the polypeptide.

The invention also provides a method for treating or preventing a disorder associated with decreased expression or activity of PROAP, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising a substantially  
10 purified polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:1-19 and fragments thereof, in conjunction with a suitable pharmaceutical carrier.

The invention also provides a method for treating or preventing a disorder associated with increased expression or activity of PROAP, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of a polypeptide having an amino acid  
15 sequence selected from the group consisting of SEQ ID NO:1-19 and fragments thereof.

#### BRIEF DESCRIPTION OF THE FIGURES AND TABLES

Figures 1A and 1B show the amino acid sequence alignment between PROAP-1 (Incyte ID number 1342011; SEQ ID NO:1) and mouse npdcf-1 (GI 452276; SEQ ID NO:39).

20 Figures 2A and 2B show the amino acid sequence alignment between PROAP-2 (Incyte ID number 1880041; SEQ ID NO:2) and human EB1 (GI 998357; SEQ ID NO:40).

Figure 3 shows the amino acid sequence alignment between PROAP-3 (Incyte ID number 3201881; SEQ ID NO:3) and mouse serum deprivation response (sdr) protein (GI 455719; SEQ ID NO:41).

25 Figure 4 shows the amino acid sequence alignment between PROAP-13 (Incyte ID number 1438978; SEQ ID NO:13) and human dim1p homolog (GI 2565275; SEQ ID NO:42).

Figures 5A and 5B show the amino acid sequence alignment between PROAP-14 (Incyte ID number 2024773; SEQ ID NO:14) and FAS-associated factor from Drosophila melanogaster (GI 3688609; SEQ ID NO:43).

30 Figure 6 shows the amino acid sequence alignment between PROAP-15 (Incyte ID number 3869790; SEQ ID NO:15) and cell death activator CIDE-B from Mus musculus (GI 3114594; SEQ ID NO:44).

The above alignments were produced using the multisequence alignment program of LASERGENE software (DNASTAR, Madison WI).

35 Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs),

clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding PROAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods and algorithms used for identification of PROAP.

5           Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

10           Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding PROAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze PROAP, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

15           Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

20           It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

25           Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing  
30           the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"PROAP" refers to the amino acid sequences of substantially purified PROAP obtained from

any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PROAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PROAP either by directly interacting with PROAP or by acting on components of the biological pathway in which PROAP participates.

An "allelic variant" is an alternative form of the gene encoding PROAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PROAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PROAP or a polypeptide with at least one functional characteristic of PROAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PROAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PROAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PROAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PROAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein



molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

5       The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PROAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PROAP either by directly interacting with PROAP or by acting on components of the biological pathway in which PROAP participates.

10       The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind PROAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the  
15 translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

20       The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

25       The term "antisense" refers to any composition containing a nucleic acid sequence which is complementary to the "sense" strand of a specific nucleic acid sequence. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" or "minus" can refer to  
30 the antisense strand, and the designation "positive" or "plus" can refer to the sense strand.

      The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic PROAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PROAP or fragments of PROAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
30	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
35	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala

	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
5	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
10	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

---

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,  
 15 (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a  
 20 polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide  
 25 from which it was derived.

A "fragment" is a unique portion of PROAP or the polynucleotide encoding PROAP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment  
 30 used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain  
 35 defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:20-38 comprises a region of unique polynucleotide sequence that

specifically identifies SEQ ID NO:20-38, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:20-38 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:20-38 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:20-38 and the region of  
5 SEQ ID NO:20-38 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-19 is encoded by a fragment of SEQ ID NO:20-38. A fragment of SEQ ID NO:1-19 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-19. For example, a fragment of SEQ ID NO:1-19 is useful as an immunogenic peptide  
10 for the development of antibodies that specifically recognize SEQ ID NO:1-19. The precise length of a fragment of SEQ ID NO:1-19 and the region of SEQ ID NO:1-19 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity  
15 or complete similarity. The word "identity" may substitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced  
20 stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be  
25 tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer  
30 to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default

parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length

supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be

used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

- 5           The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

          "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific  
10   hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive  
15   conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

- 20           Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and  
25   conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

- High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS,  
30   for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular

circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

5       The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate  
10   to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression  
15   of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

20       The term "modulate" refers to a change in the activity of PROAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PROAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or  
25   synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably  
linked to a coding sequence if the promoter affects the transcription or expression of the coding  
30   sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition.



PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding PROAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the

selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding PROAP, or fragments thereof, or PROAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

5           A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

          "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells,  
10       trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

          "Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is  
15       selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

20           A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or  
25       greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the  
30       reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The

presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## 10 THE INVENTION

The invention is based on the discovery of new human proliferation and apoptosis related proteins (PROAP), the polynucleotides encoding PROAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, immunological, and reproductive disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding PROAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each PROAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The Incyte clones in column 5 were used to assemble the consensus nucleotide sequence of each PROAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods used to identify each polypeptide through sequence homology and protein motifs.

As shown in Figures 1A and 1B, PROAP-1 has chemical and structural similarity with mouse npdcf-1 (GI 452276; SEQ ID NO:39). In particular, PROAP-1 and npdcf-1 share 66% identity and have similar isoelectric points (7.5 and 7.2, respectively). As shown in Figures 2A and 2B, PROAP-2 has chemical and structural similarity with human EB1 (GI 998357; SEQ ID NO:40). In particular, PROAP-2 and EB1 share 64% identity and have similar isoelectric points (5.3 and 4.9, respectively). As shown in Figure 3, PROAP-3 has chemical and structural similarity with mouse serum deprivation response (sdr) protein (GI 455719; SEQ ID NO:41). In particular, PROAP-3 is 86% identical to sdr

from residue M1 through V239 on sdr. As shown in Figure 4, PROAP-13 has chemical and structural similarity with human dim1p homolog (GI 2565275; SEQ ID NO:42). In particular, PROAP-13 and Dim1p share 36% identity. As shown in Figures 5A and 5B, PROAP-14 has chemical and structural similarity with Fly FAS-associated factor (FFAF) from *D. melanogaster* (GI 3688609; SEQ ID NO:43). In particular, PROAP-14 and FFAF share 40% identity. As shown in Figure 6, PROAP-15 has chemical and structural similarity with cell death activator CIDE-B from *M. musculus* (GI 3114594; SEQ ID NO:44). In particular, PROAP-15 and CIDE-B share 83% identity.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding PROAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:20-38 and to distinguish between SEQ ID NO:20-38 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express PROAP as a fraction of total tissues expressing PROAP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing PROAP as a fraction of total tissues expressing PROAP. Column 5 lists the vectors used to subclone each cDNA library. Of particular note is the expression of SEQ ID NO:20 in reproductive, nervous, and cardiovascular tissues, of SEQ ID NO:21 in nervous tissue, of SEQ ID NO:22 in reproductive and gastrointestinal tissues, of SEQ ID NO:28, which is detected exclusively in a cDNA library derived from tibia meniscus tissue, of SEQ ID NO:30, which is detected exclusively in a cDNA library derived from diseased liver, of SEQ ID NO:32 in brain tumor-associated tissues, of SEQ ID NO:33 in tumors of the breast and brain, and of SEQ ID NO:34 in tumors of the breast and testicle.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding PROAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses PROAP variants. A preferred PROAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PROAP amino acid sequence, and which contains at least one functional or structural characteristic of PROAP.

The invention also encompasses polynucleotides which encode PROAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:20-38, which encodes PROAP.

The invention also encompasses a variant of a polynucleotide sequence encoding PROAP. In

particular, such a variant polynucleotide sequence will have at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PROAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID  
5 NO:20-38 which has at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:20-38. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PROAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the  
10 genetic code, a multitude of polynucleotide sequences encoding PROAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the  
15 polynucleotide sequence of naturally occurring PROAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PROAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PROAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding  
20 PROAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PROAP and its derivatives without altering the encoded amino acid  
25 sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PROAP and PROAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell  
30 systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PROAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:20-38 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and

S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PROAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed

using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

5           When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

10           Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate  
15           software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

          In another embodiment of the invention, polynucleotide sequences or fragments thereof  
20           which encode PROAP may be cloned in recombinant DNA molecules that direct expression of PROAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PROAP.

          The nucleotide sequences of the present invention can be engineered using methods generally  
25           known in the art in order to alter PROAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction  
30           sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

          In another embodiment, sequences encoding PROAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PROAP itself or a fragment thereof may be synthesized using chemical methods. For



example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of PROAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active PROAP, the nucleotide sequences encoding PROAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PROAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PROAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PROAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PROAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences

encoding PROAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PROAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PROAP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PROAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PROAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of PROAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PROAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PROAP. Transcription of sequences encoding PROAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PROAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain  
5 infective virus which expresses PROAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of  
10 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression  
15 of PROAP in cell lines is preferred. For example, sequences encoding PROAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to  
20 confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine  
25 phosphoribosyltransferase genes, for use in *tk* and *apv* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,  
30 Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate

luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PROAP is inserted within a marker gene sequence, transformed cells containing sequences encoding PROAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PROAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PROAP and that express PROAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PROAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PROAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PROAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PROAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega

(Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PROAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PROAP may be designed to contain signal sequences which direct secretion of PROAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PROAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PROAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PROAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PROAP encoding sequence and the heterologous protein sequence, so that PROAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10).

A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PROAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

Fragments of PROAP may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of PROAP may be synthesized separately and then combined to produce the full length molecule.

### THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PROAP and proliferation and apoptosis related proteins. In addition, the expression of PROAP is closely associated with cancer, inflammation, and proliferating, reproductive, and developmental tissues. Therefore, PROAP appears to play a role in cell proliferative, immunological, and reproductive disorders. In the treatment of disorders associated with increased PROAP expression or activity, it is desirable to decrease the expression or activity of PROAP. In the treatment of disorders associated with decreased PROAP expression or activity, it is desirable to increase the expression or activity of PROAP.

Therefore, in one embodiment, PROAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PROAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-

candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, a complication of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a reproductive disorder such as disorders of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PROAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PROAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified PROAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PROAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PROAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PROAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of PROAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PROAP. Examples of such disorders include, but are not limited to, those cell proliferative, immunological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PROAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PROAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide

encoding PROAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PROAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PROAP may be produced using methods which are generally known in the art. In particular, purified PROAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PROAP. Antibodies to PROAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PROAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PROAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of PROAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PROAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J.



Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PROAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PROAP may also be generated. For example, such fragments include, but are not limited to, F(ab'), fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PROAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PROAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PROAP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of PROAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PROAP epitopes, represents the average affinity, or avidity, of the antibodies

for PROAP. The  $K_d$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PROAP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_d$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the PROAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody  
5 preparations with  $K_d$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PROAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

10 The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PROAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and  
15 guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PROAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding PROAP may be used in situations in which it would be  
20 desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding PROAP. Thus, complementary molecules or fragments may be used to modulate PROAP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding  
25 PROAP.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides  
30 encoding PROAP. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding PROAP can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding PROAP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules

until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding PROAP. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PROAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding PROAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of PROAP, antibodies to PROAP, and mimetics, agonists, antagonists, or inhibitors of PROAP. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

5        Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose,  
10    hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar  
15    solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of  
20    gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

25        Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily  
30    injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating,

5 dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any  
10 or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of PROAP, such labeling would include amount, frequency, and method of administration.

15 Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs.

20 An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PROAP or fragments thereof, antibodies of PROAP, and agonists, antagonists or inhibitors of  
25 PROAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions  
30 which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind PROAP may be used for the diagnosis of disorders characterized by expression of PROAP, or in assays to monitor patients being treated with PROAP or agonists, antagonists, or inhibitors of PROAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics.

Diagnostic assays for PROAP include methods which utilize the antibody and a label to detect PROAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PROAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PROAP expression. Normal or standard values for PROAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to PROAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PROAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PROAP may be used

for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PROAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess  
5 expression of PROAP, and to monitor regulation of PROAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PROAP or closely related molecules may be used to identify nucleic acid sequences which encode PROAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a  
10 conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PROAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PROAP encoding sequences. The hybridization probes of the subject  
15 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:20-38 or from genomic sequences including promoters, enhancers, and introns of the PROAP gene.

Means for producing specific hybridization probes for DNAs encoding PROAP include the cloning of polynucleotide sequences encoding PROAP or PROAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may  
20 be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PROAP may be used for the diagnosis of disorders  
25 associated with expression of PROAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in  
30 particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis,



autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,

5 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner

10 syndrome, a complication of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a reproductive disorder such as disorders of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids,

15 autoimmune disorders, ectopic pregnancies, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PROAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in

20 PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PROAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PROAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

25 sequences encoding PROAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PROAP in the

30 sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PROAP, a normal or standard profile for expression is established. This may be accomplished by

combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PROAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PROAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PROAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding PROAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of PROAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray

can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

5           Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

10           In another embodiment of the invention, nucleic acid sequences encoding PROAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single  
15   chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

          Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the  
20   Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PROAP on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

25           In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides  
30   valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention

may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PROAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PROAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PROAP, or fragments thereof, and washed. Bound PROAP is then detected by methods well known in the art. Purified PROAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PROAP specifically compete with a test compound for binding PROAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PROAP.

In additional embodiments, the nucleotide sequences which encode PROAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. [Atty Docket No. PF-0619 P, filed October 20, 1998] U.S. Ser. No. 60/118,559, U.S. Ser. No. [Atty Docket No. PF-0670 P, filed February 11, 1999], and U.S. Ser. No. 60/154,336 are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), or pINCY (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

### II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit

from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence

alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:20-38. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact

within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding PROAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation/trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### V. Extension of PROAP Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:20-38 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN



quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:20-38 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

## 30 VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:20-38 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06

software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech).

- 5 An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16  
10 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

## VII. Microarrays

- 15 A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After  
20 hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

- Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise  
25 the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking  
30 followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

### VIII. Complementary Polynucleotides

Sequences complementary to the PROAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PROAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same  
5 procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PROAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PROAP-encoding  
10 transcript.

### IX. Expression of PROAP

Expression and purification of PROAP is achieved using bacterial or virus-based expression systems. For expression of PROAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA  
15 transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PROAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PROAP in eukaryotic cells is achieved by infecting  
20 insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PROAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to  
25 infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PROAP is synthesized as a fusion protein with, e.g., glutathione  
30 S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from

PROAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, 5 supra, ch. 10 and 16). Purified PROAP obtained by these methods can be used directly in the following activity assay.

#### **X. Demonstration of PROAP Activity**

An assay for PROAP activity measures cell proliferation as the amount of newly initiated DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding PROAP is 10 transfected into quiescent 3T3 cultured cells using methods well known in the art. The transiently transfected cells are then incubated in the presence of [<sup>3</sup>H]thymidine, a radioactive DNA precursor. . Where applicable, varying amounts of PROAP ligand are added to the transfected cells. Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized 15 DNA.

An alternative assay for PROAP activity measures the induction of apoptosis when PROAP is expressed at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies, Gaithersburg, MD) 20 and pCR 3.1 (Invitrogen, Carlsbad, CA, both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, preferably of endothelial or hematopoietic origin, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a 25 reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP) (Clontech, Palo Alto, CA), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate their apoptotic state. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or 30 coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane

composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface.

Alternatively, PROAP activity may be measured by the induction of growth arrest when PROAP is expressed at physiologically elevated levels in transformed mammalian cell lines. PROAP cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression, and these constructs are stably transfected into a transformed cell line, such as NIH 3T6 or C6, using methods known in the art. An additional plasmid, containing sequences which encode a selectable marker, such as hygromycin resistance, are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Cells expressing PROAP are compared with control cells, either non-transfected or transfected with vector alone, for characteristics associated with growth arrest. Such characteristics can include, but are not limited to, a reduction in [<sup>3</sup>H]-thymidine incorporation into newly synthesized DNA, lower doubling and generation times, and decreased culture saturation density.

Alternatively, an assay for PROAP activity uses radiolabeled nucleotides, such as [<sup>32</sup>P]ATP, to measure either the incorporation of radiolabel into DNA during DNA synthesis, or fragmentation of DNA that accompanies apoptosis. Mammalian cells are transfected with plasmid containing cDNA encoding PROAP by methods well known in the art. Cells are then incubated with radiolabeled nucleotide for various lengths of time. Chromosomal DNA is collected, and radioactivity detected using a scintillation counter. Incorporation of radiolabel into chromosomal DNA is proportional to the degree of stimulation of the cell cycle. To determine if PROAP promotes apoptosis, chromosomal DNA is collected as above, and analyzed using polyacrylamide gel electrophoresis, by methods well known in the art. Fragmentation of DNA is quantified by comparison to untransfected control cells, and is proportional to the apoptotic activity of PROAP.

## **XI. Functional Assays**

PROAP function is assessed by expressing the sequences encoding PROAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein

provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PROAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PROAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PROAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

## **XII. Production of PROAP Specific Antibodies**

PROAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PROAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-

KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PROAP activity by, for example, binding the peptide or PROAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

### **XIII. Purification of Naturally Occurring PROAP Using Specific Antibodies**

5 Naturally occurring or recombinant PROAP is substantially purified by immunoaffinity chromatography using antibodies specific for PROAP. An immunoaffinity column is constructed by covalently coupling anti-PROAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

10 Media containing PROAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PROAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PROAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PROAP is collected.

### **15 XIV. Identification of Molecules Which Interact with PROAP**

PROAP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PROAP, washed, and any wells with labeled PROAP complex are assayed. Data obtained using  
20 different concentrations of PROAP are used to calculate values for the number, affinity, and association of PROAP with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be  
25 understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	20	1342011	COLNTUT03	1291596H1 (BRAINOT11), 485081X18 (HNT2RAT01), 671427H1 (CRBLNOT01), 1352964T6 (LATRTUT02), 1342011H1 (COLNTUT03), 1444182R1 (THRYNOT03), 1444182F1 (THRYNOT03)
2	21	1880041	LEUKNOT03	3470287H1 (BRAIDIT01), 1832158R6 (BRAINON01), 2288712H1 (BRAINON01), 1384536F1 (BRAITUT08), 1880041H1 (LEUKNOT03)
3	22	3201881	PENCNOT02	3201881H1 (PENCNOT02), 2520087F6 (BRAITUT21), 352438X15 (LVENNOT01)
4	23	939000	CERVNOT01	110900F1 (PITUNOT01), 548840F1 (BEPINOT01), 939000H1 (CERVNOT01), 939000X12 (CERVNOT01), 1271295F6 (TESTTUT02), 2122589F6 (BRSTNOT07), 3618041H1 (EPIPNOT01), SXAA02479D1, SXAA01641D1, SXAA01631D1, SAOA02385F1
5	24	2125677	BRSTNOT07	368085R1 (SYNORAT01), 392816H1 (TMLR2DT01), 518806R6 (MMLR1DT01), 1271911H1 (TESTTUT02), 1822315X314D1 (GBLATUT01), 1858290F6 (PROSNOT18), 2125677H1 (BRSTNOT07), 2293815H1 (BRAINON01), 2573443R6 (HIPOAZT01), 2764062H1 (BRSTNOT12), 2832044T6 (TLYMNOT03), 3428001H1 (BRSTNOR01), 3687264H1 (HEAANOT01), 3765525H1 (BRSTNOT24), 4590195H1 (MASTTUT01)
6	25	2603810	LUNGTUT07	013535R1 (THP1PLB01), 267329R1 (HNT2NOT01), 1453513F1 (PENITUT01), 1556582F6 (BLADTUT04), 2603810H1 (LUNGTUT07)
7	26	2715761	THYRNOT09	2715761H1 (THYRNOT09), 2993353F6 (KIDNFET02), SBLA03719F1



Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
8	27	3255641	OVARTUN01	516590H1 (MMLR1DT01), 1921460R6 (BRSTTUT01), 2824323F6 (ADRETUT06), 3255641H1 (OVARTUN01), 3255641R6 (OVARTUN01), SBXA03995D1
9	28	3620391	MENTNOT01	1556171H1 (BLADTUT04), 3620391H1 (MENTNOT01)
10	29	3969860	PROSTUT10	3969860H1 (PROSTUT10), 4275630F6 (PROSTMT01), 4275630T6 (PROSTMT01), 4403647F6 (PROSDIT01)
11	30	4286006	LIVRDIR01	4286006F6 (LIVRDIR01), 4286006H1 (LIVRDIR01)
12	31	4325626	TLYMUNT01	841543R1 (PROSTUT05), 841543X53 (PROSTUT05), 1752767F6 (LIVRTUT01), 2994209T6 (KIDNFET02), 3053308H1 (LNODNOT08), 4325626H1 (TLYMUNT01), 5209052H1 (BRAFN0T02)
13	32	1438978	PANCNOT08	834140H1 (PROSNOT07), 1438978F6 (PANCNOT08), 4074639H1 (PANCNOT19)
14	33	2024773	KERANOT02	782716R1 (MYOMNOT01), 980866R1 (TONGTUT), 1995464T6 (BRSTTUT03), 2027443H1 (KERANOT02), 2106331R6 (BRAITUT03), 3333150H1 (BRAIFET01)
15	34	3869790	BMARNOT03	359792R6 (SYNORAB01), 1535116T1 (SPLNNOT04), 2587946F6 (BRAITUT22), 3869790H1 (BMARNOT03)
16	35	001273	U937NOT01	001273H1 (U937NOT01), 1528039F1 (UCMCL5T01), 1526245F6 (UCMCL5T01), 899008R6 (BRSTTUT03), 022308F1 (ADENINE01)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
17	36	411831	BRSTNOT01	411831H1 (BRSTNOT01), 1232212F1 (LUNGFET03), 1997123R6 (BRSTTUT03), 001732H1 (U937NOT01), 414405T6 (BRSTNOT01), 781412R1 (MYOMNOT01), SADC11822F1
18	37	1520835	BLADTUT04	1419118F6 (KIDNNOT09), 1520835F1 and 1520835H1 (BLADTUT04), 1529102F6 (UCMCL5T01), 3842242F6 (DENDNOT01)
19	38	1902803	OVARNOT07	180897F1 (PLACNOB01), 491345H1 (HNT2AGT01), 927993R1 (BRAINOT04), 1902803H1 (OVARNOT07), 4217475H1 (ADRENOT15)

Table 2

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
1	334	S122 T60 S192 S203 S204 S218 S89 S118 S226	N190		Mouse npdcf-1 (g452276)	BLAST
2	281	S120 S44 S180 S245 S284 S285 T295 S143 T225 T232			Human EB1 (g998357)	BLAST
3	237	S16 T33 S149 S172 S190 Y119	N14 N25 N31 N147		Mouse serum deprivation response protein (sdr) (g455719)	BLAST
4	941	T542 T858 T30 T55 T76 T153 S159 T198 T249 T266 S300 T432 S653 S750 T29 S315 T322 T357 S372 S403 T462 S493 S572 T674 S681 S783 S853 T867 Y131 Y658	N74 N196		TPR protein (Zer1p) (g1209391)	MOTIFS BLAST

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
5	918	T19 T94 S469 T2 S44 T82 S107 T120 S257 T276 T399 S475 S579 S605 S708 S715 S785 T790 S814 S835 S841 S8 S22 S29 S60 S198 S251 S285 T374 S556 S589 S602 T634 S697 T843 T872 S897	N116	Polyadenylate binding (PABP) protein domain: P87-D126 P139-G185 R492-I568 HECT (ubiquitin transferase) domain: S605-V918	Drosophila hyper-plastic discs (HYD) protein (g2673887)	MOTIFS BLAST PFAM BLOCKS
6	324	S140 S191 S273 T287 S226		Mitochondrial energy transfer protein signature: P141-L149 Transmembrane domains: V306-I324 A33-R53	Similar to human growth arrest inducible gene product (g1707054)	MOTIFS BLAST HMM
7	185	T72 T73 T132 T21 T160 T174 S35 S95			APC10 (Anaphase promoting complex) (g3402334)	MOTIFS BLAST

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
8	445	T281 S32 S118 S135 S177 S416 T418 T81 T186 T203 S262 S302 T335 T346	N300 N414	Rhodopsin-like GPCR fingerprint: F282-L306 Transmembrane domains: I147-Y166 S357-Y373	Mitogen- induced protein (g2290726)	MOTIFS BLAST PRINTS HMM
9	73	T55 T15 S25 S28 T50	N34		Cyclin E (g1262821)	MOTIFS BLAST
10	288	T159 T161 S190 S228 S245 S56 S117 S120 S143 S190 T240	N226	SPRY domain: E132-W153 C148-M273 C3HC4 zinc finger: C11-Q39	RET finger protein-like 1, long variant (g3417312)	MOTIFS BLAST PFAM BLOCKS
11	98	T61 S22 Y57 Y69 Y90	N59	SH3 domain: A46-E64	Melanoma inhibitor protein homolog (g1778171)	MOTIFS BLAST BLOCKS PRINTS

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
12	549	S139 T313 T351 T61 T460 S484 T511 S73 S90 S91 T152 S216 T282 T315 S346 S446 Y99		Probable rabGAP domain: A98-T315	TRE oncogene product (g37330)	MOTIFS BLAST PFAM
13	95	T9 S10 S20 T48			Human dim1p homolog (g2565275)	BLAST
14	445	T14 T24 T109 S142 T213 T244 S275 Y297 S300 S355 S361 S372 S393 T425 T432	N269 N284 N370		Fly FAS-associated factor (FFAF) (g3688609)	BLAST
15	219	T46 T55 T82 T199	N18		Cell death activator CIDE-B (g3114594)	BLAST
16	439	T27 T32 S75 S123 S347 T381 T404 T263 Y231 Y294		Signal peptide: M1-A28	p52 apoptotic protein (g259942)	MOTIFS BLAST HMM

Table 2 (cont.)

Polypeptide SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Identification	Analytical Methods
17	526	S383 S470 S69 S78 S137 T273 T274 S342 S432 T453 S231 T285 T290 S342 T360 T407 S423 S436 S460 S508	N217 N229	bZIP transcription factor: K384-R398 Cyclin cell cycle division protein: A224-I250 Signal peptide: M1-S25	cyclin ania-6a g5453421 [Mus musculus]	MOTIFS BLAST BLOCKS HMM
18	298	T63 S93 S165 S212 S220 S6 T44 S133 T203 T251		C3HC4 type Zn finger: C267-A276 apoptosis inhibitor: R90-L155	putative apoptosis inhibitor (g2957175)	MOTIFS PFAM PROFILESAN BLAST
19	249	S57 S119 T134 S150 T167 S205 S52 S125 T230 Y121		PHD finger: P196-E245	candidate tumor suppressor (g2829208)	MOTIFS BLAST PFAM

Table 3

Polynucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
20	518-568		Cell Proliferation (0.660) Inflammation/Trauma (0.270)	pINCY
21	613-693		Cell Proliferation (0.560)	pINCY
22	949-984		Cell Proliferation (0.560)	pINCY
23	811-855 1297-1341	Reproductive (0.287) Nervous (0.181) Hematopoietic/Immune (0.138)	Cancer (0.487) Inflammation (0.250) Cell Proliferation (0.181)	PSPORT1
24	275-322 1955-1999	Reproductive (0.279) Nervous (0.174) Hematopoietic/Immune (0.116)	Cancer (0.419) Inflammation (0.267) Cell Proliferation (0.174)	pINCY
25	322-351	Reproductive (0.306) Cardiovascular (0.105) Hematopoietic/Immune (0.105)	Cancer (0.484) Inflammation (0.290) Cell Proliferation (0.234)	pINCY
26	658-702	Reproductive (0.444) Developmental (0.111) Hematopoietic/Immune (0.111)	Cancer (0.500) Inflammation (0.333) Cell Proliferation (0.167)	pINCY
27	172-216 604-648	Reproductive (0.256) Nervous (0.186) Hematopoietic/Immune (0.163)	Cancer (0.349) Inflammation (0.302) Trauma (0.116)	PSPORT1
28	58-102	Musculoskeletal (1.000)	Cancer (1.000)	pINCY



Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
29	217-246 433-477	Reproductive (0.455) Nervous (0.273) Cardiovascular (0.091)	Cancer (0.455) Cell Proliferation (0.182) Trauma (0.182)	pINCY
30	257-301	Gastrointestinal (1.000)	Inflammation (1.000)	pINCY
31	219-263 1569-1613	Gastrointestinal (0.245) Nervous (0.245) Reproductive (0.245)	Cancer (0.490) Inflammation (0.265) Cell Proliferation (0.143)	pINCY
32	585-629	Nervous (0.390) Reproductive (0.150)	Cancer and Cell Proliferation (0.690)	
33	381-425	Reproductive (0.310) Nervous (0.150)	Cancer and Cell Proliferation (0.650)	
34	133-177	Reproductive (0.330)	Cancer (0.440)	
35	110-154	Reproductive (0.282) Hematopoietic/Immune (0.256) Cardiovascular (0.154)	Cancer (0.462) Inflammation (0.256) Fetal (0.179)	PBLUESCRIPT
36	164-208	Reproductive (0.236) Gastrointestinal (0.181) Hematopoietic/Immune (0.153)	Cancer (0.486) Inflammation (0.264) Fetal (0.125)	PBLUESCRIPT

Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
37	272-316	Developmental (0.429) Hematopoietic/Immune (0.286) Reproductive (0.143) Urologic (0.143)	Fetal (0.571) Cancer (0.286) Inflammation (0.143)	pINCY
38	782-826	Reproductive (0.253) Nervous (0.176) Urologic (0.121)	Cancer (0.440) Inflammation (0.242) Fetal (0.231)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
20	COLNTUT03	This library was constructed using RNA isolated from colon tumor tissue obtained from the sigmoid colon of a 62-year-old Caucasian male during a sigmoidectomy and permanent colostomy. Pathology indicated invasive grade 2 adenocarcinoma. One lymph node contained metastasis with extranodal extension. Patient history included hyperlipidemia, cataract disorder, and dermatitis. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, breast cancer, and prostate cancer.
21	LEUKNOT03	This library was constructed using RNA isolated from white blood cells of a 27-year-old female with blood type A+. The donor tested negative for cytomegalovirus (CMV).
22	PENCNOT02	This library was constructed using RNA isolated from penis right corpus cavernosum tissue.
23	CERVNOT01	This library was constructed using RNA isolated from uterine cervical tissue of a 35-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated mild chronic cervicitis. Family history included atherosclerotic coronary artery disease and type II diabetes.
24	BRSTNOT07	This library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
25	LUNGTUT07	This library was constructed using RNA isolated from lung tumor tissue removed from the upper lobe of a 50-year-old Caucasian male during segmental lung resection. Pathology indicated an invasive grade 4 squamous cell adenocarcinoma. Patient history included tobacco use. Family history included skin cancer.
26	THYRN0T09	This library was constructed using RNA isolated from diseased thyroid tissue removed from an 18-year-old Caucasian female during a unilateral thyroid lobectomy and regional lymph node excision. Pathology indicated adenomatous goiter associated with a follicular adenoma of the thyroid. Family history included thyroid cancer.
27	OVARTUN01	This normalized library was constructed from 5.36 million independent clones obtained from an ovarian tumor library. RNA was isolated from tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a single ovary, and inguinal hernia repair. Pathology indicated metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary and a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular ovarian cyst, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, hyperlipidemia, and primary tuberculous complex. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9928) and Bonaldo et al. (Genome Research (1996) 6:791).

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
28	MENTNOT01	This library was constructed using RNA isolated from left tibial meniscus tissue removed from a 16-year-old Caucasian male during a partial left tibial osteotomy with free skin graft. Pathology for the associated tumor indicated metastatic alveolar rhabdomyosarcoma. Patient history included an abnormality of the red blood cells. Family history included osteoarthritis.
29	PROSTUT10	This library was constructed using RNA isolated from prostatic tumor tissue removed from a 66-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated an adenocarcinoma (Gleason grade 2+3) and adenofibromatous hyperplasia. The patient presented with elevated prostate specific antigen (PSA). Family history included prostate cancer and secondary bone cancer.
30	LIVRDIR01	This library was constructed using RNA isolated from diseased liver tissue removed from a 63-year-old Caucasian female during a liver transplant. Patient history included primary biliary cirrhosis. Serology was positive for anti-mitochondrial antibody.
31	TLYMUNT01	This library was constructed using RNA isolated from resting allogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male.
32	PANCNOT08	This library was constructed using RNA isolated from pancreatic tissue removed from a 65-year-old Caucasian female during radical subtotal pancreatectomy. Pathology for the associated tumor tissue indicated an invasive grade 2 adenocarcinoma. Patient history included type II diabetes, osteoarthritis, cardiovascular disease, benign neoplasm in the large bowel, and a cataract.
33	KERANOT02	This library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
34	BMARNOT03	This library was constructed using RNA isolated from the left tibial bone marrow tissue of a 16-year-old Caucasian male during a partial left tibial osteotomy with free skin graft. Patient history included an abnormality of the red blood cells. Previous surgeries included bone and bone marrow biopsy, and soft tissue excision.
35	U937NOT01	This library was constructed at Stratagene (STR937207), using RNA isolated from the U937 monocyte-like cell line. This line (ATCC CRL1593) was established from malignant cells obtained from the pleural effusion of a 37-year-old Caucasian male with diffuse histiocytic lymphoma.
36	BRSTNOT01	This library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.
37	BLADTUT04	This library was constructed using RNA isolated from bladder tumor tissue removed from a 60-year-old Caucasian male during a radical cystectomy, prostatectomy, and vasectomy. Pathology indicated grade 3 transitional cell carcinoma in the left bladder wall. Carcinoma in-situ was identified in the dome and trigone. Family history included type I diabetes, a malignant neoplasm of the stomach, atherosclerotic coronary artery disease, and an acute myocardial infarction.
38	OVARNOT07	This library was constructed using RNA isolated from left ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. The tissue was associated with multiple follicular cysts, endometrium in a weakly proliferative phase, and chronic cervicitis of the cervix with squamous metaplasia. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value= 1.0E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS and PRINTS databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and Probability value= 1.0E-3 or less
PFAM	A Hidden Markov Models-based application useful for protein family search.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits, depending on individual protein families

Table 5 cont.

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Score= 4.0 or greater
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <u>supra</u> ; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	



What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, and fragments thereof.
2. A substantially purified variant having at least 90% amino acid sequence identity to the amino acid sequence of claim 1.
3. An isolated and purified polynucleotide encoding the polypeptide of claim 1.
4. An isolated and purified polynucleotide variant having at least 90% polynucleotide sequence identity to the polynucleotide of claim 3.
5. An isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide of claim 3.
6. An isolated and purified polynucleotide having a sequence which is complementary to the polynucleotide of claim 3.
7. A method for detecting a polynucleotide, the method comprising the steps of:
  - (a) hybridizing the polynucleotide of claim 6 to at least one nucleic acid in a sample, thereby forming a hybridization complex; and
  - (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of the polynucleotide in the sample.
8. The method of claim 7 further comprising amplifying the polynucleotide prior to hybridization.
9. An isolated and purified polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID

NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, and fragments thereof.

10. An isolated and purified polynucleotide variant having at least 90% polynucleotide  
5 sequence identity to the polynucleotide of claim 9.

11. An isolated and purified polynucleotide having a sequence which is complementary  
to the polynucleotide of claim 9.

10 12. An expression vector comprising at least a fragment of the polynucleotide of claim 3.

13. A host cell comprising the expression vector of claim 12.

14. A method for producing a polypeptide, the method comprising the steps of:  
15 a) culturing the host cell of claim 13 under conditions suitable for the  
expression of the polypeptide; and  
b) recovering the polypeptide from the host cell culture.

15. A pharmaceutical composition comprising the polypeptide of claim 1 in conjunction  
20 with a suitable pharmaceutical carrier.

16. A purified antibody which specifically binds to the polypeptide of claim 1.

17. A purified agonist of the polypeptide of claim 1.

25

18. A purified antagonist of the polypeptide of claim 1.

19. A method for treating or preventing a disorder associated with decreased expression  
or activity of PROAP, the method comprising administering to a subject in need of such treatment an  
30 effective amount of the pharmaceutical composition of claim 15.

20. A method for treating or preventing a disorder associated with increased expression  
or activity of PROAP, the method comprising administering to a subject in need of such treatment an  
effective amount of the antagonist of claim 18.

1	M	S	R	T	M	A	R	T	R	P	G	Q	L	G	-	-	R	V	T	G	A	G	W	G	S	A	A	V	C	1342011	
1	M	A	T	P	V	P	P	S	P	R	H	L	R	L	R	L	L	S	G	-	-	-	-	-	-	-	-	L	I	GI452276	
29	R	G	R	A	L	R	G	R	E	P	A	L	P	S	A	S	F	P	D	V	A	A	C	P	G	S	L	D	C	A	1342011
25	L	G	A	A	L	N	G	-	-	-	-	A	T	A	R	R	P	D	A	T	T	C	P	G	S	L	D	C	A	GI452276	
59	L	K	R	R	A	R	C	P	P	G	A	H	A	C	G	P	C	L	Q	P	F	Q	E	D	Q	Q	G	L	C	V	1342011
50	L	K	R	R	A	K	C	P	P	G	A	H	A	C	G	P	C	L	Q	S	F	Q	E	D	Q	R	G	F	C	V	GI452276
89	P	R	M	R	R	P	P	G	G	R	P	Q	P	R	L	E	D	E	I	D	F	L	A	Q	E	L	A	-	-	1342011	
80	P	R	K	H	L	S	S	G	E	G	L	P	Q	P	R	L	E	E	I	D	S	L	A	Q	E	L	A	L	K	GI452276	
117	R	K	E	S	G	H	S	-	-	-	T	P	P	L	P	K	D	R	Q	R	L	P	E	P	A	-	T	L	G	F	1342011
110	E	K	E	A	G	H	S	R	L	T	A	Q	P	L	L	E	R	A	Q	K	L	L	E	P	A	A	T	L	G	F	GI452276
143	S	A	R	G	Q	G	L	E	L	G	L	P	S	T	P	G	T	P	T	P	H	T	S	L	G	S	P	V	1342011		
140	S	Q	W	G	Q	R	L	E	P	G	L	P	S	T	H	G	T	S	S	P	I	P	H	T	S	L	S	R	A	GI452276	
173	S	S	D	P	V	H	M	S	P	L	E	P	R	G	G	Q	G	D	G	L	A	L	V	L	I	L	A	F	C	V	1342011
170	S	S	G	P	V	Q	M	S	P	L	E	P	Q	G	R	H	G	N	G	L	T	L	V	L	I	L	A	F	C	L	GI452276

FIGURE 1A

203	A	G	A	A	L	S	V	A	S	L	C	W	C	R	L	Q	R	E	I	R	L	T	Q	K	A	D	Y	A	-	1342011	
200	A	S	A	A	L	A	V	A	A	L	C	W	C	R	L	Q	R	E	I	R	L	T	Q	K	A	D	Y	A	A	GI452276	
232	T	A	K	A	P	G	S	P	A	A	P	R	I	S	P	G	D	Q	R	L	A	Q	S	A	E	M	Y	H	Y	Q	1342011
230	T	A	K	G	P	T	S	P	S	T	P	R	I	S	P	G	D	Q	R	L	A	H	S	A	E	M	Y	H	Y	Q	GI452276
262	H	Q	R	Q	Q	M	L	C	L	E	R	H	K	E	P	P	K	E	L	D	T	A	S	S	D	E	E	N	E	D	1342011
260	H	Q	R	Q	Q	M	L	C	L	E	R	H	K	E	P	P	K	E	L	E	S	A	S	S	D	E	E	N	E	D	GI452276
292	G	D	F	T	V	Y	E	C	P	G	L	A	P	T	G	E	M	E	V	R	N	P	L	F	D	H	A	A	L	S	1342011
290	G	D	F	T	V	Y	E	C	P	G	L	A	P	T	G	E	M	E	V	R	N	P	L	F	D	H	S	T	L	S	GI452276
322	A	P	L	P	A	P	S	S	P	P	A	L	P																	1342011	
320	A	P	V	P	G	P	H	S	L	P	P	L	Q																	GI452276	

FIGURE 1B

1	M	A	V	N	V	Y	S	T	S	V	T	S	E	N	L	S	R	H	D	M	L	A	W	V	N	D	S	L	H	L	1880041
1	M	A	V	N	V	Y	S	T	S	V	T	S	D	N	L	S	R	H	D	M	L	A	W	I	N	E	S	L	Q	L	GI 998357
31	N	Y	T	K	I	E	Q	L	C	S	G	A	A	Y	C	Q	F	M	D	M	L	F	P	G	C	V	H	L	R	K	1880041
31	N	L	T	K	I	E	Q	L	C	S	G	A	A	Y	C	Q	F	M	D	M	L	F	P	G	S	I	A	L	K	K	GI 998357
61	V	K	F	Q	A	K	L	E	H	E	Y	I	H	N	F	K	V	L	Q	A	A	F	K	K	M	G	V	D	K	I	1880041
61	V	K	F	Q	A	K	L	E	H	E	Y	I	Q	N	F	K	I	L	Q	A	G	F	K	R	M	G	V	D	K	I	GI 998357
91	I	P	V	E	K	L	V	K	G	K	F	Q	D	N	F	E	F	I	Q	W	F	K	K	F	F	D	A	N	Y	D	1880041
91	I	P	V	D	K	L	V	K	G	K	F	Q	D	N	F	E	F	V	Q	W	F	K	K	F	F	D	A	N	Y	D	GI 998357
121	G	K	D	Y	N	P	L	L	A	R	Q	G	Q	D	V	A	P	P	N	P	G	D	Q	I	F	N	K	S	K	1880041	
121	G	K	D	Y	D	P	V	A	A	R	Q	G	Q	E	T	A	V	A	P	S	L	V	A	P	A	L	N	K	P	K	GI 998357
151	K	L	I	G	T	A	V	P	Q	R	T	S	P	T	G	P	K	N	M	Q	T	S	G	R	L	S	N	V	A	P	1880041
151	K	P	L	T	S	S	A	A	P	Q	R	P	I	S	T	Q	R	T	A	A	P	K	-	-	-	-	-	-	A	G	GI 998357

FIGURE 2A

181	P	C	I	L	R	K	N	P	P	S	A	R	N	G	G	H	E	T	D	A	Q	I	L	E	L	N	Q	Q	L	V	1880041
177	P	G	V	V	R	K	N	P	-	-	-	-	G	V	G	N	G	D	D	E	A	A	E	L	M	Q	Q	V	N	GI 998357	
211	D	L	K	L	T	V	D	G	L	E	K	E	R	D	F	Y	F	S	K	L	R	D	I	E	L	I	C	Q	E	H	1880041
202	V	L	K	L	T	V	E	D	L	E	K	E	R	D	F	Y	F	G	K	L	R	N	I	E	L	I	C	Q	E	N	GI 998357
241	E	S	E	N	S	P	V	I	S	G	I	I	G	I	L	Y	A	T	E	E	G	F	A	P	P	E	D	D	E	I	1880041
232	E	G	E	N	D	P	V	L	Q	R	I	V	D	I	L	Y	A	T	D	E	G	F	V	I	P	D	E	G	G	-	GI 998357
271	E	E	H	Q	Q	E	D	Q	D	E	Y																			1880041	
261	-	-	-	-	P	Q	E	E	Q	E	E	Y																		GI 998357	

FIGURE 2B

1	M	G	E	D	A	A	Q	A	E	K	F	Q	H	P	G	S	D	M	R	Q	E	K	P	S	S	P	S	P	M	P	3201881
1	M	G	E	D	A	A	Q	A	E	K	F	Q	H	P	N	T	D	M	L	Q	E	K	P	S	S	P	S	P	M	P	GI 455719
31	S	S	T	P	S	P	S	L	N	L	G	N	T	E	E	A	I	R	D	N	S	Q	V	N	A	V	T	V	L	T	3201881
31	S	S	T	P	S	P	S	L	N	L	G	S	T	E	E	A	I	R	D	N	S	Q	V	N	A	V	T	V	H	T	GI 455719
61	L	L	D	K	L	V	N	M	L	D	A	V	Q	E	N	Q	H	K	M	E	Q	R	Q	I	S	L	E	G	S	V	3201881
61	L	L	D	K	L	V	N	M	L	D	A	V	R	E	N	Q	H	N	M	E	Q	R	Q	I	N	L	E	G	S	V	GI 455719
91	K	G	I	Q	N	D	L	T	K	L	S	K	Y	Q	A	S	T	S	N	T	V	S	K	L	L	E	K	S	R	K	3201881
91	K	G	I	Q	N	D	L	T	K	L	S	K	Y	Q	A	S	T	S	N	T	V	S	K	L	L	E	K	S	R	K	GI 455719
121	V	S	A	H	T	R	A	V	K	E	R	M	D	R	Q	C	A	Q	V	K	R	L	E	N	N	H	A	Q	L	L	3201881
121	V	S	A	H	T	R	A	V	R	E	R	L	E	R	Q	C	V	Q	V	K	R	L	E	N	N	H	A	Q	L	L	GI 455719
151	R	R	N	H	F	K	V	L	I	F	Q	E	E	N	E	I	P	A	S	V	F	V	K	Q	P	V	S	G	A	V	3201881
151	R	R	N	H	F	K	V	L	I	F	Q	E	E	S	E	I	P	A	S	V	F	V	K	E	P	V	P	S	A	A	GI 455719
181	E	G	K	E	E	L	P	D	E	N	K	S	L	E	E	T	L	H	T	V	D	L	S	S	D	D	D	L	P	H	3201881
181	E	G	K	E	E	L	A	D	E	N	K	S	L	E	E	T	L	H	N	V	D	L	S	S	D	D	E	L	P	R	GI 455719
211	D	E	E	A	L	E	D	S	A	E	E	K	V	G	R	S	P	R	G	R	E	I	K	R	S	-	-	-	R	P	3201881
211	D	E	E	A	L	E	D	S	A	E	E	K	M	E	-	E	S	R	A	E	K	I	K	R	S	S	L	K	K	V	GI 455719

FIGURE 3

6/9

1	M	S	F	L	L	P	K	L	T	S	K	K	E	V	D	Q	A	I	K	S	T	A	E	K	V	L	V	L	R	F	1438978
1	M	S	Y	M	L	P	H	L	H	N	G	W	Q	V	D	Q	A	I	L	S	E	E	D	R	V	V	V	I	R	F	GI 2565275
31	G	R	D	E	D	P	V	C	L	Q	L	D	I	L	S	K	T	S	S	D	L	S	K	M	A	A	I	Y	L	1438978	
31	G	H	D	W	D	P	T	C	M	K	M	D	E	V	L	Y	S	I	A	E	K	V	K	N	F	A	V	I	Y	L	GI 2565275
61	V	D	V	D	Q	T	A	V	Y	T	Q	Y	F	D	I	S	Y	I	P	S	T	V	-	F	F	F	N	G	Q	H	1438978
61	V	D	I	T	E	V	P	D	F	N	K	M	Y	E	L	-	Y	D	P	C	T	V	M	F	F	F	R	N	K	H	GI 2565275
90	M	K	V	D	Y	G																								1438978	
90	I	M	I	D	L	G	T	G	N	N	N	K	I	N	W	A	M	E	D	K	Q	E	M	V	D	I	I	E	T	V	GI 2565275
95																															1438978
120	Y	R	G	A	R	K	G	R	G	L	V	S	P	K	D	Y	S	T	K	Y	R	Y									GI 2565275

FIGURE 4



1 M A A P E E R D L T Q E Q T E K L L Q F Q D D L T G I E S M D 2024773  
 1 M - - - E A D G L T N E Q T E K V L Q F Q D D L T G I E D M N GI 3688609  
  
 31 Q C R H T L E Q H N W N I E A A V Q D R L N E Q E G V P S V 2024773  
 28 V C R D V L I R H Q W D L E V A F Q E Q L N I R E G R P T M GI 3688609  
  
 61 F N P P P S - - - - - R P L Q - - - - V N T A D H R 2024773  
 58 F A A S T D V R A P A V L N D R F L Q Q V F S A N M P G G R GI 3688609  
  
 78 I Y S Y V V S R P Q P R G L L G W G Y Y L I M L P F R F T Y 2024773  
 88 T V S R V P S G P V P R S F T G I I G Y V I N F V F Q Y - F GI 3688609  
  
 108 Y T I L D I F R F A L R F I R P D P R S R - V T D P V G D I 2024773  
 117 Y S T L T S I V S A F V N L G G N E A R L V T D P L G D V GI 3688609  
  
 137 V S F M H S F E E K Y G R A H P V F Y Q G T Y S Q A L N D A 2024773  
 147 M K F I R E Y Y E R Y - P E H P V F Y Q G T Y A Q A L N D A GI 3688609  
  
 167 K R E L R F L L V Y L H G D - - D H Q D S D E F C R N T L C 2024773  
 176 K Q E L R F L I V Y L H K D P A K N P D V E S F C R N T L S GI 3688609  
  
 195 A P E V I S L I N T R M L F W A C S T N K P E G Y R V S Q A 2024773  
 206 A R S V I D Y I N T H T L L W G C D V A T P E G Y R V M Q S GI 3688609  
  
 225 L R E N T Y P F L A M I M L K D R R M T V V G R L E G L I Q 2024773  
 236 I T V R S Y P T M V M I S L R A N R M M I V G R F E G D C T GI 3688609

FIGURE 5A

255 P D D L I N Q L T F I M D A N Q T Y L V S E R L E R E E R N 2024773  
 266 P E E L L R R L Q S V T N A N E V W L S Q A R A D R L E R N GI 3688609  
  
 285 Q T Q V L R Q Q Q D E A Y L A S L R A D Q E K E R K K R E E 2024773  
 296 F T Q T L R Q Q Q D E A Y E Q S L L A D E E K E R Q R Q R E GI 3688609  
  
 315 R E R K R R K E E V Q Q Q K L A E E R R R Q N L Q E E K E 2024773  
 326 R D A V R Q A E E A V E Q A R R D V E L R K E E I A R Q K I GI 3688609  
  
 345 R K L E C L P P E P S P D D P E S V K I I F K L P N D S R V 2024773  
 356 E L A T L V P S E P A A D A V G A I A V V F K L P S G T R L GI 3688609  
  
 375 E R R F H F S Q S L T V I H D F L F S L K E S P E K F Q I E 2024773  
 386 E R R F N Q T D S V L D V Y H Y L F C H P D S P D E F E I T GI 3688609  
  
 405 A N F P R R V L P C I P S E E W P N P P - - - - - T 2024773  
 416 T N F P K R V L F S K A N L D A A G E T G T A K E T L T K T GI 3688609  
  
 426 L Q E A G L S H T E V L F V Q D L T D E 2024773  
 446 L Q A V G L K N R E L L F V N D L - E A GI 3688609

FIGURE 5B

9/9

1	M E Y L S A L N P S D L L R S V S N I S S E F G R R V W T S	3869790
1	M E Y L S A F N P N G L L R S V S T V S S E L S R R V W N S	GI 3114594
31	A P P P Q R P F R V C D H K R T I R K G L T A A T R Q E L I	3869790
31	A P P P Q R P F R V C D H K R T V R K G L T A A S L Q E L I	GI 3114594
61	A K A L E T L L L N G V L T L V L E E D G T A V D S E D F F	3869790
61	D K V L E T L L L R G V L T L V L E E D G T A V D S E D F F	GI 3114594
91	Q L L E D D T C L M V L Q S G Q S W S P T R S G V L S Y G I	3869790
91	Q L L E D D T C L M V L E Q G Q S W S P - K S G M L S Y G I	GI 3114594
121	G R E R P K H S K D I A R F T F D V Y K Q N P R D L F G S I	3869790
120	G R E K P K H S K D I A R I T F D V Y K Q N P R D L F G S I	GI 3114594
151	N V K A T F Y G L Y S M S C D F Q G L G P K K V L R E L L F	3869790
150	N V K A T F Y G L Y S M S C D F Q G V G P K R V L R E L L F	GI 3114594
181	W T S T L L Q G L G H M L L G I S S T L R H A V E G A E Q W	3869790
180	G T S S Q L Q G L G H M L L G I S S T L R H V V E G A D R W	GI 3114594
211	Q Q K G - - R L H S Y	3869790
210	Q W H G Q R H L H S	GI 3114594

FIGURE 6

## SEQUENCE LISTING

&lt;110&gt; INCYTE PHARMACEUTICALS, INC.

TANG, Y. Tom

YUE, Henry

HILLMAN, Jennifer L.

GUEGLER, Karl J.

CORLEY, Neil C.

LAL, Preeti

AZIMZAI, Yalda

BAUGHN, Mariah R.

JUNMING, Yang

SHIH, Leo L.

&lt;120&gt; PROLIFERATION AND APOPTOSIS RELATED PROTEINS

&lt;130&gt; PF-0619 PCT

&lt;140&gt; To Be Assigned

&lt;141&gt; Herewith

<150> 09/175,737; unassigned; 60/118,559; 09/249,740; unassigned;  
60/154,336<151> 1998-10-20; 1998-10-20; 1999-02-04; 1999-04-11; 1999-04-11;  
1999-04-22

&lt;160&gt; 44

&lt;170&gt; PERL Program

&lt;210&gt; 1

&lt;211&gt; 334

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1342011CD1

&lt;400&gt; 1

Met	Ser	Arg	Thr	Met	Ala	Arg	Thr	Arg	Pro	Gly	Gln	Leu	Gly	Arg
1				5					10					15
Val	Thr	Gly	Ala	Gly	Gly	Trp	Gly	Ser	Ala	Ala	Val	Cys	Arg	Gly
				20					25					30
Arg	Ala	Leu	Arg	Gly	Arg	Glu	Pro	Ala	Leu	Pro	Ser	Ala	Ser	Phe
				35					40					45
Pro	Asp	Val	Ala	Ala	Cys	Pro	Gly	Ser	Leu	Asp	Cys	Ala	Leu	Lys
				50					55					60
Arg	Arg	Ala	Arg	Cys	Pro	Pro	Gly	Ala	His	Ala	Cys	Gly	Pro	Cys
				65					70					75
Leu	Gln	Pro	Phe	Gln	Glu	Asp	Gln	Gln	Gly	Leu	Cys	Val	Pro	Arg
				80					85					90
Met	Arg	Arg	Pro	Pro	Gly	Gly	Gly	Arg	Pro	Gln	Pro	Arg	Leu	Glu
				95					100					105

```

Asp Glu Ile Asp Phe Leu Ala Gln Glu Leu Ala Arg Lys Glu Ser
    110                      115                      120
Gly His Ser Thr Pro Pro Leu Pro Lys Asp Arg Gln Arg Leu Pro
    125                      130                      135
Glu Pro Ala Thr Leu Gly Phe Ser Ala Arg Gly Gln Gly Leu Glu
    140                      145                      150
Leu Gly Leu Pro Ser Thr Pro Gly Thr Pro Thr Pro Thr Pro His
    155                      160                      165
Thr Ser Leu Gly Ser Pro Val Ser Ser Asp Pro Val His Met Ser
    170                      175                      180
Pro Leu Glu Pro Arg Gly Gly Gln Gly Asp Gly Leu Ala Leu Val
    185                      190                      195
Leu Ile Leu Ala Phe Cys Val Ala Gly Ala Ala Ala Leu Ser Val
    200                      205                      210
Ala Ser Leu Cys Trp Cys Arg Leu Gln Arg Glu Ile Arg Leu Thr
    215                      220                      225
Gln Lys Ala Asp Tyr Ala Thr Ala Lys Ala Pro Gly Ser Pro Ala
    230                      235                      240
Ala Pro Arg Ile Ser Pro Gly Asp Gln Arg Leu Ala Gln Ser Ala
    245                      250                      255
Glu Met Tyr His Tyr Gln His Gln Arg Gln Gln Met Leu Cys Leu
    260                      265                      270
Glu Arg His Lys Glu Pro Pro Lys Glu Leu Asp Thr Ala Ser Ser
    275                      280                      285
Asp Glu Glu Asn Glu Asp Gly Asp Phe Thr Val Tyr Glu Cys Pro
    290                      295                      300
Gly Leu Ala Pro Thr Gly Glu Met Glu Val Arg Asn Pro Leu Phe
    305                      310                      315
Asp His Ala Ala Leu Ser Ala Pro Leu Pro Ala Pro Ser Ser Pro
    320                      325                      330
Pro Ala Leu Pro

```

&lt;210&gt; 2

&lt;211&gt; 281

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1880041CD1

&lt;400&gt; 2

```

Met Ala Val Asn Val Tyr Ser Thr Ser Val Thr Ser Glu Asn Leu
    1                      5                      10                      15
Ser Arg His Asp Met Leu Ala Trp Val Asn Asp Ser Leu His Leu
    20                      25                      30
Asn Tyr Thr Lys Ile Glu Gln Leu Cys Ser Gly Ala Ala Tyr Cys
    35                      40                      45
Gln Phe Met Asp Met Leu Phe Pro Gly Cys Val His Leu Arg Lys
    50                      55                      60
Val Lys Phe Gln Ala Lys Leu Glu His Glu Tyr Ile His Asn Phe
    65                      70                      75
Lys Val Leu Gln Ala Ala Phe Lys Lys Met Gly Val Asp Lys Ile
    80                      85                      90

```

```

Ile Pro Val Glu Lys Leu Val Lys Gly Lys Phe Gln Asp Asn Phe
      95      100      105
Glu Phe Ile Gln Trp Phe Lys Lys Phe Phe Asp Ala Asn Tyr Asp
      110      115      120
Gly Lys Asp Tyr Asn Pro Leu Leu Ala Arg Gln Gly Gln Asp Val
      125      130      135
Ala Pro Pro Pro Asn Pro Gly Asp Gln Ile Phe Asn Lys Ser Lys
      140      145      150
Lys Leu Ile Gly Thr Ala Val Pro Gln Arg Thr Ser Pro Thr Gly
      155      160      165
Pro Lys Asn Met Gln Thr Ser Gly Arg Leu Ser Asn Val Ala Pro
      170      175      180
Pro Cys Ile Leu Arg Lys Asn Pro Pro Ser Ala Arg Asn Gly Gly
      185      190      195
His Glu Thr Asp Ala Gln Ile Leu Glu Leu Asn Gln Gln Leu Val
      200      205      210
Asp Leu Lys Leu Thr Val Asp Gly Leu Glu Lys Glu Arg Asp Phe
      215      220      225
Tyr Phe Ser Lys Leu Arg Asp Ile Glu Leu Ile Cys Gln Glu His
      230      235      240
Glu Ser Glu Asn Ser Pro Val Ile Ser Gly Ile Ile Gly Ile Leu
      245      250      255
Tyr Ala Thr Glu Glu Gly Phe Ala Pro Pro Glu Asp Asp Glu Ile
      260      265      270
Glu Glu His Gln Gln Glu Asp Gln Asp Glu Tyr
      275      280

```

&lt;210&gt; 3

&lt;211&gt; 237

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3201881CD1

&lt;400&gt; 3

```

Met Gly Glu Asp Ala Ala Gln Ala Glu Lys Phe Gln His Pro Gly
  1      5      10      15
Ser Asp Met Arg Gln Glu Lys Pro Ser Ser Pro Ser Pro Met Pro
      20      25      30
Ser Ser Thr Pro Ser Pro Ser Leu Asn Leu Gly Asn Thr Glu Glu
      35      40      45
Ala Ile Arg Asp Asn Ser Gln Val Asn Ala Val Thr Val Leu Thr
      50      55      60
Leu Leu Asp Lys Leu Val Asn Met Leu Asp Ala Val Gln Glu Asn
      65      70      75
Gln His Lys Met Glu Gln Arg Gln Ile Ser Leu Glu Gly Ser Val
      80      85      90
Lys Gly Ile Gln Asn Asp Leu Thr Lys Leu Ser Lys Tyr Gln Ala
      95      100      105
Ser Thr Ser Asn Thr Val Ser Lys Leu Leu Glu Lys Ser Arg Lys
      110      115      120
Val Ser Ala His Thr Arg Ala Val Lys Glu Arg Met Asp Arg Gln

```

	125		130		135
Cys Ala Gln Val	Lys Arg Leu Glu Asn	Asn His Ala Gln Leu Leu			
	140		145		150
Arg Arg Asn His	Phe Lys Val Leu Ile	Phe Gln Glu Glu Asn Glu			
	155		160		165
Ile Pro Ala Ser	Val Phe Val Lys Gln	Pro Val Ser Gly Ala Val			
	170		175		180
Glu Gly Lys Glu	Glu Leu Pro Asp Glu	Asn Lys Ser Leu Glu Glu			
	185		190		195
Thr Leu His Thr	Val Asp Leu Ser Ser	Asp Asp Asp Leu Pro His			
	200		205		210
Asp Glu Glu Ala	Leu Glu Asp Ser Ala	Glu Glu Lys Val Gly Arg			
	215		220		225
Ser Pro Arg Gly	Arg Glu Ile Lys Arg	Ser Arg Pro			
	230		235		

&lt;210&gt; 4

&lt;211&gt; 941

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 939000CD1

&lt;400&gt; 4

Met Asn Lys Lys Lys Lys Pro Phe Leu Gly Met Pro Ala Pro Leu		
1	5	10 15
Gly Tyr Val Pro Gly Leu Gly Arg Gly Ala Thr Gly Phe Thr Thr		
	20	25 30
Arg Ser Asp Ile Gly Pro Ala Arg Asp Ala Asn Asp Pro Val Asp		
	35	40 45
Asp Arg His Ala Pro Pro Gly Lys Arg Thr Val Gly Asp Gln Met		
	50	55 60
Lys Lys Asn Gln Ala Ala Asp Asp Asp Asp Glu Asp Leu Asn Asp		
	65	70 75
Thr Asn Tyr Asp Glu Phe Asn Gly Tyr Ala Gly Ser Leu Phe Ser		
	80	85 90
Ser Gly Pro Tyr Glu Lys Asp Asp Glu Glu Ala Asp Ala Ile Tyr		
	95	100 105
Ala Ala Leu Asp Lys Arg Met Asp Glu Arg Arg Lys Glu Arg Arg		
	110	115 120
Glu Gln Arg Glu Lys Glu Glu Ile Glu Lys Tyr Arg Met Glu Arg		
	125	130 135
Pro Lys Ile Gln Gln Gln Phe Ser Asp Leu Lys Arg Lys Leu Ala		
	140	145 150
Glu Val Thr Glu Glu Glu Trp Leu Ser Ile Pro Glu Val Gly Asp		
	155	160 165
Ala Arg Asn Lys Arg Gln Arg Asn Pro Arg Tyr Glu Lys Leu Thr		
	170	175 180
Pro Val Pro Asp Ser Phe Phe Ala Lys His Leu Gln Thr Gly Glu		
	185	190 195
Asn His Thr Ser Val Asp Pro Arg Gln Thr Gln Phe Gly Gly Leu		
	200	205 210

Asn Thr Pro Tyr	Pro Gly Gly Leu Asn Thr	Pro Tyr Pro Gly Gly	
215	220	225	
Met Thr Pro Gly	Leu Met Thr Pro Gly Thr Gly Glu Leu Asp Met		
230	235	240	
Arg Lys Ile Gly	Gln Ala Arg Asn Thr Leu Met Asp Met Arg Leu		
245	250	255	
Ser Gln Val Ser	Asp Ser Val Ser Gly Gln Thr Val Val Asp Pro		
260	265	270	
Lys Gly Tyr Leu	Thr Asp Leu Asn Ser Met Ile Pro Thr His Gly		
275	280	285	
Gly Asp Ile Asn	Asp Ile Lys Lys Ala Arg Leu Leu Leu Lys Ser		
290	295	300	
Val Arg Glu Thr	Asn Pro His His Pro Pro Ala Trp Ile Ala Ser		
305	310	315	
Ala Arg Leu Glu	Glu Val Thr Gly Lys Leu Gln Val Ala Arg Asn		
320	325	330	
Leu Ile Met Lys	Gly Thr Glu Met Cys Pro Lys Ser Glu Asp Val		
335	340	345	
Trp Leu Glu Ala	Ala Arg Leu Gln Pro Gly Asp Thr Ala Lys Ala		
350	355	360	
Val Val Ala Gln	Ala Val Arg His Leu Pro Gln Ser Val Arg Ile		
365	370	375	
Tyr Ile Arg Ala	Ala Glu Leu Glu Thr Asp Ile Arg Ala Lys Lys		
380	385	390	
Arg Val Leu Arg	Lys Ala Leu Glu His Val Pro Asn Ser Val Arg		
395	400	405	
Leu Trp Lys Ala	Ala Val Glu Leu Glu Glu Pro Glu Asp Ala Arg		
410	415	420	
Ile Met Leu Ser	Arg Ala Val Glu Cys Cys Pro Thr Ser Val Glu		
425	430	435	
Leu Trp Leu Ala	Leu Ala Arg Leu Glu Thr Tyr Glu Asn Ala Arg		
440	445	450	
Lys Val Leu Asn	Lys Ala Arg Glu Asn Ile Pro Thr Asp Arg His		
455	460	465	
Ile Trp Ile Thr	Ala Ala Lys Leu Glu Glu Ala Asn Gly Asn Thr		
470	475	480	
Gln Met Val Glu	Lys Ile Ile Asp Arg Ala Ile Thr Ser Leu Arg		
485	490	495	
Ala Asn Gly Val	Glu Ile Asn Arg Glu Gln Trp Ile Gln Asp Ala		
500	505	510	
Glu Glu Cys Asp	Arg Ala Gly Ser Val Ala Thr Cys Gln Ala Val		
515	520	525	
Met Arg Ala Val	Ile Gly Ile Gly Ile Glu Glu Glu Asp Arg Lys		
530	535	540	
His Thr Trp Met	Glu Asp Ala Asp Ser Cys Val Ala His Asn Ala		
545	550	555	
Leu Glu Cys Ala	Arg Ala Ile Tyr Ala Tyr Ala Leu Gln Val Phe		
560	565	570	
Pro Ser Lys Lys	Ser Val Trp Leu Arg Ala Ala Tyr Phe Glu Lys		
575	580	585	
Asn His Gly Thr	Arg Glu Ser Leu Glu Ala Leu Leu Gln Arg Ala		
590	595	600	
Val Ala His Cys	Pro Lys Ala Glu Val Leu Trp Leu Met Gly Ala		
605	610	615	
Lys Ser Lys Trp	Leu Ala Gly Asp Val Pro Ala Ala Arg Ser Ile		



	620		625		630
Leu Ala Leu Ala	Phe Gln Ala Asn Pro	Asn Ser Glu Glu Ile Trp			
	635		640		645
Leu Ala Ala Val	Lys Leu Glu Ser Glu	Asn Asp Glu Tyr Glu Arg			
	650		655		660
Ala Arg Arg Leu	Leu Ala Lys Ala Arg	Ser Ser Ala Pro Thr Ala			
	665		670		675
Arg Val Phe Met	Lys Ser Val Lys Leu	Glu Trp Val Gln Asp Asn			
	680		685		690
Ile Arg Ala Ala	Gln Asp Leu Cys Glu	Glu Ala Leu Arg His Tyr			
	695		700		705
Glu Asp Phe Pro	Lys Leu Trp Met Met	Lys Gly Gln Ile Glu Glu			
	710		715		720
Gln Lys Glu Met	Met Glu Lys Ala Arg	Glu Ala Tyr Asn Gln Gly			
	725		730		735
Leu Lys Lys Cys	Pro His Ser Thr Pro	Leu Trp Leu Leu Leu Ser			
	740		745		750
Arg Leu Glu Glu	Lys Ile Gly Gln Leu	Thr Arg Ala Arg Ala Ile			
	755		760		765
Leu Glu Lys Ser	Arg Leu Lys Asn Pro	Lys Asn Pro Gly Leu Trp			
	770		775		780
Leu Glu Ser Val	Arg Leu Glu Tyr Arg	Ala Gly Leu Lys Asn Ile			
	785		790		795
Ala Asn Thr Leu	Met Ala Lys Ala Leu	Gln Glu Cys Pro Asn Ser			
	800		805		810
Gly Ile Leu Trp	Ser Glu Ala Ile Phe	Leu Glu Ala Arg Pro Gln			
	815		820		825
Arg Arg Thr Lys	Ser Val Asp Ala Leu	Lys Lys Cys Glu His Asp			
	830		835		840
Pro His Val Leu	Leu Ala Val Ala Lys	Leu Phe Trp Ser Gln Arg			
	845		850		855
Lys Ile Thr Lys	Ala Arg Glu Trp Phe	His Arg Thr Val Lys Ile			
	860		865		870
Asp Ser Asp Leu	Gly Asp Ala Trp Ala	Phe Phe Tyr Lys Phe Glu			
	875		880		885
Leu Gln His Gly	Thr Glu Glu Gln Gln	Glu Glu Val Arg Lys Arg			
	890		895		900
Cys Glu Ser Ala	Glu Pro Arg His Gly	Glu Leu Trp Cys Ala Val			
	905		910		915
Ser Lys Asp Ile	Ala Asn Trp Gln Lys	Lys Ile Gly Asp Ile Leu			
	920		925		930
Arg Leu Val Ala	Gly Arg Ile Lys Asn	Thr Phe			
	935		940		

&lt;210&gt; 5

&lt;211&gt; 918

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2125677CD1

&lt;400&gt; 5

Met Thr Ala Arg	Glu Glu Ala Ser Leu Arg Thr Leu Glu Gly Arg	
1	5	10 15
Arg Arg Ala Thr	Leu Leu Ser Ala Arg Gln Gly Met Met Ser Ala	
	20	25 30
Arg Gly Asp Phe	Leu Asn Tyr Ala Leu Ser Leu Met Arg Ser His	
	35	40 45
Asn Asp Glu His	Ser Asp Val Leu Pro Val Leu Asp Val Cys Ser	
	50	55 60
Leu Lys His Val	Ala Tyr Val Phe Gln Ala Leu Ile Tyr Trp Ile	
	65	70 75
Lys Ala Met Asn	Gln Gln Thr Thr Leu Asp Thr Pro Gln Leu Glu	
	80	85 90
Arg Lys Arg Thr	Arg Glu Leu Leu Glu Leu Gly Ile Asp Asn Glu	
	95	100 105
Asp Ser Glu His	Glu Asn Asp Asp Asp Thr Asn Gln Ser Ala Thr	
	110	115 120
Leu Asn Asp Lys	Asp Asp Asp Ser Leu Pro Ala Glu Thr Gly Gln	
	125	130 135
Asn His Pro Phe	Phe Arg Arg Ser Asp Ser Met Thr Phe Leu Gly	
	140	145 150
Cys Ile Pro Pro	Asn Pro Phe Glu Val Pro Leu Ala Glu Ala Ile	
	155	160 165
Pro Leu Ala Asp	Gln Pro His Leu Leu Gln Pro Asn Ala Arg Lys	
	170	175 180
Glu Asp Leu Phe	Gly Arg Pro Ser Gln Gly Leu Tyr Ser Ser Ser	
	185	190 195
Ala Ser Ser Gly	Lys Cys Leu Met Glu Val Thr Val Asp Arg Asn	
	200	205 210
Cys Leu Glu Val	Leu Pro Thr Lys Met Ser Tyr Ala Ala Asn Leu	
	215	220 225
Lys Asn Val Met	Asn Met Gln Asn Arg Gln Lys Lys Glu Gly Glu	
	230	235 240
Glu Gln Pro Val	Leu Pro Glu Glu Thr Glu Ser Ser Lys Pro Gly	
	245	250 255
Pro Ser Ala His	Asp Leu Ala Ala Gln Leu Lys Ser Ser Leu Leu	
	260	265 270
Ala Glu Ile Gly	Leu Thr Glu Ser Glu Gly Pro Pro Leu Thr Ser	
	275	280 285
Phe Arg Pro Gln	Cys Ser Phe Met Gly Met Val Ile Ser His Asp	
	290	295 300
Met Leu Leu Gly	Arg Trp Arg Leu Ser Leu Glu Leu Phe Gly Arg	
	305	310 315
Val Phe Met Glu	Asp Val Gly Ala Glu Pro Gly Ser Ile Leu Thr	
	320	325 330
Glu Leu Gly Gly	Phe Glu Val Lys Glu Ser Lys Phe Arg Arg Glu	
	335	340 345
Met Glu Lys Leu	Arg Asn Gln Gln Ser Arg Asp Leu Ser Leu Glu	
	350	355 360
Val Lys Val Asp	Arg Asp Arg Asp Leu Leu Ile Gln Gln Thr Met	
	365	370 375
Arg Gln Leu Asn	Asn His Phe Gly Arg Arg Cys Ala Thr Thr Pro	
	380	385 390
Met Ala Val His	Arg Val Lys Val Thr Phe Lys Asp Glu Pro Gly	
	395	400 405

Glu Gly Ser Gly	Val Ala Arg Ser Phe Tyr Thr Ala Ile Ala Gln	
	410	415 420
Ala Phe Leu Ser	Asn Glu Lys Leu Pro Asn Leu Glu Cys Ile Gln	
	425	430 435
Asn Ala Asn Lys	Gly Thr His Thr Ser Leu Met Gln Arg Leu Arg	
	440	445 450
Asn Arg Gly Glu	Arg Asp Arg Glu Arg Glu Arg Glu Arg Glu Met	
	455	460 465
Arg Arg Ser Ser	Gly Leu Arg Ala Gly Ser Arg Arg Asp Arg Asp	
	470	475 480
Arg Asp Phe Arg	Arg Gln Leu Ser Ile Asp Thr Arg Pro Phe Arg	
	485	490 495
Pro Ala Ser Glu	Gly Asn Pro Ser Asp Asp Pro Glu Pro Leu Pro	
	500	505 510
Ala His Arg Gln	Ala Leu Gly Glu Arg Leu Tyr Pro Arg Val Gln	
	515	520 525
Ala Met Gln Pro	Ala Phe Ala Ser Lys Ile Thr Gly Met Leu Leu	
	530	535 540
Glu Leu Ser Pro	Ala Gln Leu Leu Leu Leu Leu Ala Ser Glu Asp	
	545	550 555
Ser Leu Arg Ala	Arg Val Asp Glu Ala Met Glu Leu Ile Ile Ala	
	560	565 570
His Gly Arg Glu	Asn Gly Ala Asp Ser Ile Leu Asp Leu Gly Leu	
	575	580 585
Val Asp Ser Ser	Glu Lys Val Gln Gln Glu Asn Arg Lys Arg His	
	590	595 600
Gly Ser Ser Arg	Ser Val Val Asp Met Asp Leu Asp Asp Thr Asp	
	605	610 615
Asp Gly Asp Asp	Asn Ala Pro Leu Phe Tyr Gln Pro Gly Lys Arg	
	620	625 630
Gly Phe Tyr Thr	Pro Arg Pro Gly Lys Asn Thr Glu Ala Arg Leu	
	635	640 645
Asn Cys Phe Arg	Asn Ile Gly Arg Ile Leu Gly Leu Cys Leu Leu	
	650	655 660
Gln Asn Glu Leu	Cys Pro Ile Thr Leu Asn Arg His Val Ile Lys	
	665	670 675
Val Leu Leu Gly	Arg Lys Val Asn Trp His Asp Phe Ala Phe Phe	
	680	685 690
Asp Pro Val Met	Tyr Glu Ser Leu Arg Gln Leu Ile Leu Ala Ser	
	695	700 705
Gln Ser Ser Asp	Ala Asp Ala Val Phe Ser Ala Met Asp Leu Ala	
	710	715 720
Phe Ala Ile Asp	Leu Cys Lys Glu Glu Gly Gly Gln Val Glu	
	725	730 735
Leu Ile Pro Asn	Gly Val Asn Ile Pro Val Thr Pro Gln Asn Val	
	740	745 750
Tyr Glu Tyr Val	Arg Lys Tyr Ala Glu His Arg Met Leu Val Val	
	755	760 765
Ala Glu Gln Pro	Leu His Ala Met Arg Lys Gly Leu Leu Asp Val	
	770	775 780
Leu Pro Lys Asn	Ser Leu Glu Asp Leu Thr Ala Glu Asp Phe Arg	
	785	790 795
Leu Leu Val Asn	Gly Cys Gly Glu Val Asn Val Gln Met Leu Ile	
	800	805 810
Ser Phe Thr Ser	Phe Asn Asp Glu Ser Gly Glu Asn Ala Glu Lys	

	815		820		825
Leu Leu Gln Phe	Lys Arg Trp Phe Trp	Ser Ile Val Glu Lys	Met		
	830		835		840
Ser Met Thr Glu	Arg Gln Asp Leu Val	Tyr Phe Trp Thr Ser	Ser		
	845		850		855
Pro Ser Leu Pro	Ala Ser Glu Glu Gly	Phe Gln Pro Met Pro	Ser		
	860		865		870
Ile Thr Ile Arg	Pro Pro Asp Asp Gln	His Leu Pro Thr Ala	Asn		
	875		880		885
Thr Cys Ile Ser	Arg Leu Tyr Val Pro	Leu Tyr Ser Ser Lys	Gln		
	890		895		900
Ile Leu Lys Gln	Lys Leu Leu Leu Ala	Ile Lys Thr Lys Asn	Phe		
	905		910		915
Gly Phe Val					

&lt;210&gt; 6

&lt;211&gt; 324

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2603810CD1

&lt;400&gt; 6

Met Gly Pro Trp Gly	Glu Pro Glu Leu Leu	Val Trp Arg Pro Glu	
1	5	10	15
Ala Val Ala Ser Glu	Pro Pro Val Pro Val	Gly Leu Glu Val Lys	
	20	25	30
Leu Gly Ala Leu Val	Leu Leu Leu Val Leu	Thr Leu Leu Cys Ser	
	35	40	45
Leu Val Pro Ile Cys	Val Leu Arg Arg Pro	Gly Ala Asn His Glu	
	50	55	60
Gly Ser Ala Ser Arg	Gln Lys Ala Leu Ser	Leu Val Ser Cys Phe	
	65	70	75
Ala Gly Gly Val Phe	Leu Ala Thr Cys Leu	Leu Asp Leu Leu Pro	
	80	85	90
Asp Tyr Leu Ala Ala	Ile Asp Glu Ala Leu	Ala Ala Leu His Val	
	95	100	105
Thr Leu Gln Phe Pro	Leu Gln Glu Phe Ile	Leu Ala Met Gly Phe	
	110	115	120
Phe Leu Val Leu Val	Met Glu Gln Ile Thr	Leu Ala Tyr Lys Glu	
	125	130	135
Gln Ser Gly Pro Ser	Pro Leu Glu Glu Thr	Arg Ala Leu Leu Gly	
	140	145	150
Thr Val Asn Gly Gly	Pro Gln His Trp His	Asp Gly Pro Gly Val	
	155	160	165
Pro Gln Ala Ser Gly	Ala Pro Ala Thr Pro	Ser Ala Leu Arg Ala	
	170	175	180
Cys Val Leu Val Phe	Ser Leu Ala Leu His	Ser Val Phe Glu Gly	
	185	190	195
Leu Ala Val Gly Leu	Gln Arg Asp Arg Ala	Arg Ala Met Glu Leu	
	200	205	210
Cys Leu Ala Leu Leu	Leu His Lys Gly Ile	Leu Ala Val Ser Leu	

	215		220		225
Ser Leu Arg Leu	Leu Gln Ser His Leu	Arg Ala Gln Val Val	Ala		
	230		235		240
Gly Cys Gly Ile	Leu Phe Ser Cys Met	Thr Pro Leu Gly Ile	Gly		
	245		250		255
Leu Gly Ala Ala	Leu Ala Glu Ser Ala	Gly Pro Leu His Gln	Leu		
	260		265		270
Ala Gln Ser Val	Leu Glu Gly Met Ala	Ala Gly Thr Phe Leu	Tyr		
	275		280		285
Ile Thr Phe Leu	Glu Ile Leu Pro Gln	Glu Leu Ala Ser Ser	Glu		
	290		295		300
Gln Arg Ile Leu	Lys Val Ile Leu Leu	Leu Ala Gly Phe Ala	Leu		
	305		310		315
Leu Thr Gly Leu	Leu Phe Ile Gln Ile				
	320				

&lt;210&gt; 7

&lt;211&gt; 185

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2715761CD1

&lt;400&gt; 7

Met Thr Thr Pro	Asn Lys Thr Pro	Pro Gly Ala Asp	Pro Lys Gln
1	5	10	15
Leu Glu Arg Thr	Gly Thr Val Arg	Glu Ile Gly Ser	Gln Ala Val
	20	25	30
Trp Ser Leu Ser	Ser Cys Lys Pro	Gly Phe Gly Val	Asp Gln Leu
	35	40	45
Arg Asp Asp Asn	Leu Glu Thr Tyr	Trp Gln Ser Asp	Gly Ser Gln
	50	55	60
Pro His Leu Val	Asn Ile Gln Phe	Arg Arg Lys Thr	Thr Val Lys
	65	70	75
Thr Leu Cys Ile	Tyr Ala Asp Tyr	Lys Ser Asp Glu	Ser Tyr Thr
	80	85	90
Pro Ser Lys Ile	Ser Val Arg Val	Gly Asn Asn Phe	His Asn Leu
	95	100	105
Gln Glu Ile Arg	Gln Leu Glu Leu	Val Glu Pro Ser	Gly Trp Ile
	110	115	120
His Val Pro Leu	Thr Asp Asn His	Lys Lys Pro Thr	Arg Thr Phe
	125	130	135
Met Ile Gln Ile	Ala Val Leu Ala	Asn His Gln Asn	Gly Arg Asp
	140	145	150
Thr His Met Arg	Gln Ile Lys Ile	Tyr Thr Pro Val	Glu Glu Ser
	155	160	165
Ser Ile Gly Lys	Phe Pro Arg Cys	Thr Thr Ile Asp	Phe Met Met
	170	175	180
Tyr Arg Ser Ile	Arg		
	185		

<210> 8  
 <211> 445  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3255641CD1

<400> 8

Met	Leu	Ala	Ser	Tyr	Gly	Leu	Ala	Tyr	Ser	Leu	Met	Lys	Phe	Phe
1				5					10					15
Thr	Gly	Pro	Met	Ser	Asp	Phe	Lys	Asn	Val	Gly	Leu	Val	Phe	Val
				20					25					30
Asn	Ser	Lys	Arg	Asp	Arg	Thr	Lys	Ala	Val	Leu	Cys	Met	Val	Val
				35					40					45
Ala	Gly	Ala	Ile	Ala	Ala	Val	Phe	His	Thr	Leu	Ile	Ala	Tyr	Ser
				50					55					60
Asp	Leu	Gly	Tyr	Tyr	Ile	Ile	Asn	Lys	Leu	His	His	Val	Asp	Glu
				65					70					75
Ser	Val	Gly	Ser	Lys	Thr	Arg	Arg	Ala	Phe	Leu	Tyr	Leu	Ala	Ala
				80					85					90
Phe	Pro	Phe	Met	Asp	Ala	Met	Ala	Trp	Thr	His	Ala	Gly	Ile	Leu
				95					100					105
Leu	Lys	His	Lys	Tyr	Ser	Phe	Leu	Val	Gly	Cys	Ala	Ser	Ile	Ser
				110					115					120
Asp	Val	Ile	Ala	Gln	Val	Val	Phe	Val	Ala	Ile	Leu	Leu	His	Ser
				125					130					135
His	Leu	Glu	Cys	Arg	Glu	Pro	Leu	Leu	Ile	Pro	Ile	Leu	Ser	Leu
				140					145					150
Tyr	Met	Gly	Ala	Leu	Val	Arg	Cys	Thr	Thr	Leu	Cys	Leu	Gly	Tyr
				155					160					165
Tyr	Lys	Asn	Ile	His	Asp	Ile	Ile	Pro	Asp	Arg	Ser	Gly	Pro	Glu
				170					175					180
Leu	Gly	Gly	Asp	Ala	Thr	Ile	Arg	Lys	Met	Leu	Ser	Phe	Trp	Trp
				185					190					195
Pro	Leu	Ala	Leu	Ile	Leu	Ala	Thr	Gln	Arg	Ile	Ser	Arg	Pro	Ile
				200					205					210
Val	Asn	Leu	Phe	Val	Ser	Arg	Asp	Leu	Gly	Gly	Ser	Ser	Ala	Ala
				215					220					225
Thr	Glu	Ala	Val	Ala	Ile	Leu	Thr	Ala	Thr	Tyr	Pro	Val	Gly	His
				230					235					240
Met	Pro	Tyr	Gly	Trp	Leu	Thr	Glu	Ile	Arg	Ala	Val	Tyr	Pro	Ala
				245					250					255
Phe	Asp	Lys	Asn	Asn	Pro	Ser	Asn	Lys	Leu	Val	Ser	Thr	Ser	Asn
				260					265					270
Thr	Val	Thr	Ala	Ala	His	Ile	Lys	Lys	Phe	Thr	Phe	Val	Cys	Met
				275					280					285
Ala	Leu	Ser	Leu	Thr	Leu	Cys	Phe	Val	Met	Phe	Trp	Thr	Pro	Asn
				290					295					300
Val	Ser	Glu	Lys	Ile	Leu	Ile	Asp	Ile	Ile	Gly	Val	Asp	Phe	Ala
				305					310					315
Phe	Ala	Glu	Leu	Cys	Val	Val	Pro	Leu	Arg	Ile	Phe	Ser	Phe	Phe
				320					325					330
Pro	Val	Pro	Val	Thr	Val	Arg	Ala	His	Leu	Thr	Gly	Trp	Leu	Met

	335		340		345
Thr Leu Lys Lys Thr Phe Val Leu Ala Pro Ser Ser Val Leu Arg					
	350		355		360
Ile Ile Val Leu Ile Ala Ser Leu Val Val Leu Pro Tyr Leu Gly					
	365		370		375
Val His Gly Ala Thr Leu Gly Val Gly Ser Leu Leu Ala Gly Phe					
	380		385		390
Val Gly Glu Ser Thr Met Val Ala Ile Ala Ala Cys Tyr Val Tyr					
	395		400		405
Arg Lys Gln Lys Lys Lys Met Glu Asn Glu Ser Ala Thr Glu Gly					
	410		415		420
Glu Asp Ser Ala Met Thr Asp Met Pro Pro Thr Glu Glu Val Thr					
	425		430		435
Asp Ile Val Glu Met Arg Glu Glu Asn Glu					
	440		445		

&lt;210&gt; 9

&lt;211&gt; 73

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3620391CD1

&lt;400&gt; 9

Met Pro Arg Glu Arg Arg Glu Arg Asp Ala Lys Glu Arg Asp Thr			
1	5	10	15
Met Lys Glu Asp Gly Gly Ala Glu Phe Ser Ala Arg Ser Arg Lys			
	20	25	30
Arg Lys Ala Asn Val Thr Val Phe Cys Arg Ile Gln Met Lys Lys			
	35	40	45
Trp Pro Lys Ser Thr Gly Arg Arg Trp Thr Ser Val Gly Ala Arg			
	50	55	60
Leu Gly Arg Met Met Gln Ser Val Gln Ala Pro Ala Pro			
	65	70	

&lt;210&gt; 10

&lt;211&gt; 288

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3969860CD1

&lt;400&gt; 10

Met Ala Ala Leu Phe Gln Glu Ala Ser Ser Cys Pro Val Cys Ser			
1	5	10	15
Asp Tyr Leu Glu Lys Pro Met Ser Leu Glu Cys Gly Cys Ala Val			
	20	25	30
Cys Leu Lys Cys Ile Asn Ser Leu Gln Lys Glu Pro His Gly Glu			
	35	40	45

```

Asp Leu Leu Cys Cys Cys Ser Ser Met Val Ser Arg Lys Asn Lys
      50      55      60
Ile Arg Arg Asn Arg Gln Leu Glu Arg Leu Ala Ser His Ile Lys
      65      70      75
Glu Leu Glu Pro Lys Leu Lys Lys Ile Leu Gln Met Asn Pro Arg
      80      85      90
Met Arg Lys Phe Gln Val Asp Met Thr Leu Asp Ala Asn Thr Ala
      95     100     105
Asn Asn Phe Leu Leu Ile Ser Asp Asp Leu Arg Ser Val Arg Ser
     110     115     120
Gly Arg Ile Arg Gln Asn Arg Gln Asp Leu Ala Glu Arg Phe Asp
     125     130     135
Val Ser Val Cys Ile Leu Gly Ser Pro Arg Phe Thr Cys Gly Arg
     140     145     150
His Cys Trp Glu Val Asp Val Gly Thr Ser Thr Glu Trp Asp Leu
     155     160     165
Gly Val Cys Arg Glu Ser Val His Arg Lys Gly Arg Ile Gln Leu
     170     175     180
Thr Thr Glu Leu Gly Phe Trp Thr Val Ser Leu Arg Asp Gly Gly
     185     190     195
Arg Leu Ser Ala Ser Thr Val Pro Leu Thr Phe Leu Phe Val Asp
     200     205     210
Arg Lys Leu Gln Arg Val Gly Ile Phe Leu Asp Met Gly Met Gln
     215     220     225
Asn Val Ser Phe Phe Asp Ala Glu Ser Gly Ser His Val Tyr Thr
     230     235     240
Phe Arg Ser Val Ser Ala Glu Glu Pro Leu Arg Pro Phe Leu Ala
     245     250     255
Pro Ser Val Pro Pro Asn Gly Asp Gln Gly Val Leu Ser Ile Cys
     260     265     270

Pro Leu Met Asn Ser Gly Thr Thr Asp Ala Pro Val Arg Pro Gly
     275     280     285
Glu Ala Lys

```

&lt;210&gt; 11

&lt;211&gt; 98

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4286006CD1

&lt;400&gt; 11

```

Met Ala Lys Phe Gly Val His Arg Ile Leu Leu Leu Ala Ile Ser
  1      5      10      15
Leu Thr Lys Cys Leu Glu Ser Thr Lys Leu Leu Ala Asp Leu Lys
      20      25      30
Lys Cys Gly Asp Leu Glu Cys Glu Ala Leu Ile Asn Arg Val Ser
      35      40      45
Ala Met Arg Asp Tyr Arg Gly Pro Asp Cys Arg Tyr Leu Asn Phe
      50      55      60
Thr Lys Gly Glu Glu Ile Ser Val Tyr Val Lys Leu Ala Gly Asp

```



	65		70		75
Arg Glu Asp Leu Trp	Ala Gly Ser Lys Gly	Lys Glu Phe Gly Tyr			
	80		85		90
Phe Pro Arg Asp Ala	Val Gln Ile				
	95				

&lt;210&gt; 12

&lt;211&gt; 549

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4325626CD1

&lt;400&gt; 12

Met Asp Val Val Glu Val Ala Gly Ser Trp	Trp Ala Gln Glu Arg
1	5 10 15
Glu Asp Ile Ile Met Lys Tyr Glu Lys Gly	His Arg Ala Gly Leu
	20 25 30
Pro Glu Asp Lys Lys Pro Lys Pro Phe Arg	Ser Tyr Asn Asn Asn
	35 40 45
Val Asp His Leu Gly Ile Val His Glu Thr	Glu Leu Pro Pro Leu
	50 55 60
Thr Ala Arg Glu Ala Lys Gln Ile Arg Arg	Glu Ile Ser Arg Lys
	65 70 75
Ser Lys Trp Val Asp Met Leu Gly Asp Trp	Glu Lys Tyr Lys Ser
	80 85 90
Ser Arg Lys Leu Ile Asp Arg Ala Tyr Lys	Gly Met Pro Met Asn
	95 100 105
Ile Arg Gly Pro Met Trp Ser Val Leu Leu	Asn Thr Glu Glu Met
	110 115 120
Lys Leu Lys Asn Pro Gly Arg Tyr Gln Ile	Met Lys Glu Lys Gly
	125 130 135
Lys Arg Ser Ser Glu His Ile Gln Arg Ile	Asp Arg Asp Val Ser
	140 145 150
Gly Thr Leu Arg Lys His Ile Phe Phe Arg	Asp Arg Tyr Gly Thr
	155 160 165
Lys Gln Arg Glu Leu Leu His Ile Leu Leu	Ala Tyr Glu Glu Tyr
	170 175 180
Asn Pro Glu Val Gly Tyr Cys Arg Asp Leu	Ser His Ile Ala Ala
	185 190 195
Leu Phe Leu Leu Tyr Leu Pro Glu Glu Asp	Ala Phe Trp Ala Leu
	200 205 210
Val Gln Leu Leu Ala Ser Glu Arg His Ser	Leu Gln Gly Phe His
	215 220 225
Ser Pro Asn Gly Gly Thr Val Gln Gly Leu	Gln Asp Gln Gln Glu
	230 235 240
His Val Val Ala Thr Ser Gln Pro Lys Thr	Met Gly His Gln Asp
	245 250 255
Lys Lys Asp Leu Cys Gly Gln Cys Ser Pro	Leu Gly Cys Leu Ile
	260 265 270
Arg Ile Leu Ile Asp Gly Ile Ser Leu Gly	Leu Thr Leu Arg Leu
	275 280 285
Trp Asp Val Tyr Leu Val Glu Gly Glu Gln	Ala Leu Met Pro Ile

	290		295		300
Thr Arg Ile Ala	Phe Lys Val Gln Gln	Lys Arg Leu Thr Lys Thr			
	305		310		315
Ser Arg Cys Gly	Pro Trp Ala Arg Phe	Cys Asn Arg Phe Val Asp			
	320		325		330
Thr Trp Ala Arg	Asp Glu Asp Thr Val	Leu Lys His Leu Arg Ala			
	335		340		345
Ser Met Lys Lys	Leu Thr Arg Lys Gln	Gly Asp Leu Pro Pro Pro			
	350		355		360
Ala Lys Pro Glu	Gln Gly Ser Ser Ala	Ser Arg Pro Val Pro Ala			
	365		370		375
Ser Arg Gly Gly	Lys Thr Leu Cys Lys	Gly Asp Arg Gln Ala Pro			
	380		385		390
Pro Gly Pro Pro	Ala Arg Phe Pro Arg	Pro Ile Trp Ser Ala Ser			
	395		400		405
Pro Pro Arg Ala	Pro Arg Ser Ser Thr	Pro Cys Pro Gly Gly Ala			
	410		415		420
Val Arg Glu Asp	Thr Tyr Pro Val Gly	Thr Gln Gly Val Pro Ser			
	425		430		435
Pro Ala Leu Ala	Gln Gly Gly Pro Gln	Gly Ser Trp Arg Phe Leu			
	440		445		450
Gln Trp Asn Ser	Met Pro Arg Leu Pro	Thr Asp Leu Asp Val Glu			
	455		460		465
Gly Pro Trp Phe	Arg His Tyr Asp Phe	Arg Gln Ser Cys Trp Val			
	470		475		480
Arg Ala Ile Ser	Gln Glu Asp Gln Leu	Ala Pro Cys Trp Gln Ala			
	485		490		495
Glu His Pro Ala	Glu Arg Val Arg Ser	Ala Phe Ala Ala Pro Ser			
	500		505		510
Thr Asp Ser Asp	Gln Gly Thr Pro Phe	Arg Ala Arg Asp Glu Gln			
	515		520		525
Pro Cys Ala Pro	Thr Ser Gly Pro Cys	Leu Cys Gly Leu His Leu			
	530		535		540
Glu Ser Ser Gln	Phe Pro Pro Gly Phe				
	545				

&lt;210&gt; 13

&lt;211&gt; 95

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1438978CD1

&lt;400&gt; 13

Met Ser Phe Leu Leu	Pro Lys Leu Thr Ser	Lys Lys Glu Val Asp
1	5	10 15
Gln Ala Ile Lys Ser	Thr Ala Glu Lys Val	Leu Val Leu Arg Phe
	20	25 30
Gly Arg Asp Glu Asp	Pro Val Cys Leu Gln	Leu Asp Asp Ile Leu
	35	40 45
Ser Lys Thr Ser Ser	Asp Leu Ser Lys Met	Ala Ala Ile Tyr Leu
	50	55 60

Val Asp Val Asp Gln Thr Ala Val Tyr Thr Gln Tyr Phe Asp Ile  
 65 70 75  
 Ser Tyr Ile Pro Ser Thr Val Phe Phe Phe Asn Gly Gln His Met  
 80 85 90  
 Lys Val Asp Tyr Gly  
 95

&lt;210&gt; 14

&lt;211&gt; 445

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2024773CD1

&lt;400&gt; 14

Met Ala Ala Pro Glu Glu Arg Asp Leu Thr Gln Glu Gln Thr Glu  
 1 5 10 15  
 Lys Leu Leu Gln Phe Gln Asp Leu Thr Gly Ile Glu Ser Met Asp  
 20 25 30  
 Gln Cys Arg His Thr Leu Glu Gln His Asn Trp Asn Ile Glu Ala  
 35 40 45  
 Ala Val Gln Asp Arg Leu Asn Glu Gln Glu Gly Val Pro Ser Val  
 50 55 60  
 Phe Asn Pro Pro Pro Ser Arg Pro Leu Gln Val Asn Thr Ala Asp  
 65 70 75  
 His Arg Ile Tyr Ser Tyr Val Val Ser Arg Pro Gln Pro Arg Gly  
 80 85 90  
 Leu Leu Gly Trp Gly Tyr Tyr Leu Ile Met Leu Pro Phe Arg Phe  
 95 100 105  
 Thr Tyr Tyr Thr Ile Leu Asp Ile Phe Arg Phe Ala Leu Arg Phe  
 110 115 120  
 Ile Arg Pro Asp Pro Arg Ser Arg Val Thr Asp Pro Val Gly Asp  
 125 130 135  
 Ile Val Ser Phe Met His Ser Phe Glu Glu Lys Tyr Gly Arg Ala  
 140 145 150  
 His Pro Val Phe Tyr Gln Gly Thr Tyr Ser Gln Ala Leu Asn Asp  
 155 160 165  
 Ala Lys Arg Glu Leu Arg Phe Leu Leu Val Tyr Leu His Gly Asp  
 170 175 180  
 Asp His Gln Asp Ser Asp Glu Phe Cys Arg Asn Thr Leu Cys Ala  
 185 190 195  
 Pro Glu Val Ile Ser Leu Ile Asn Thr Arg Met Leu Phe Trp Ala  
 200 205 210  
 Cys Ser Thr Asn Lys Pro Glu Gly Tyr Arg Val Ser Gln Ala Leu  
 215 220 225  
 Arg Glu Asn Thr Tyr Pro Phe Leu Ala Met Ile Met Leu Lys Asp  
 230 235 240  
 Arg Arg Met Thr Val Val Gly Arg Leu Glu Gly Leu Ile Gln Pro  
 245 250 255  
 Asp Asp Leu Ile Asn Gln Leu Thr Phe Ile Met Asp Ala Asn Gln  
 260 265 270  
 Thr Tyr Leu Val Ser Glu Arg Leu Glu Arg Glu Glu Arg Asn Gln  
 275 280 285

```

Thr Gln Val Leu Arg Gln Gln Gln Asp Glu Ala Tyr Leu Ala Ser
      290      295      300
Leu Arg Ala Asp Gln Glu Lys Glu Arg Lys Lys Arg Glu Glu Arg
      305      310      315
Glu Arg Lys Arg Arg Lys Glu Glu Glu Val Gln Gln Gln Lys Leu
      320      325      330
Ala Glu Glu Arg Arg Arg Gln Asn Leu Gln Glu Glu Lys Glu Arg
      335      340      345
Lys Leu Glu Cys Leu Pro Pro Glu Pro Ser Pro Asp Asp Pro Glu
      350      355      360
Ser Val Lys Ile Ile Phe Lys Leu Pro Asn Asp Ser Arg Val Glu
      365      370      375
Arg Arg Phe His Phe Ser Gln Ser Leu Thr Val Ile His Asp Phe
      380      385      390
Leu Phe Ser Leu Lys Glu Ser Pro Glu Lys Phe Gln Ile Glu Ala
      395      400      405
Asn Phe Pro Arg Arg Val Leu Pro Cys Ile Pro Ser Glu Glu Trp
      410      415      420
Pro Asn Pro Pro Thr Leu Gln Glu Ala Gly Leu Ser His Thr Glu
      425      430      435
Val Leu Phe Val Gln Asp Leu Thr Asp Glu
      440      445

```

&lt;210&gt; 15

&lt;211&gt; 219

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3869790CD1

&lt;400&gt; 15

```

Met Glu Tyr Leu Ser Ala Leu Asn Pro Ser Asp Leu Leu Arg Ser
  1      5      10      15
Val Ser Asn Ile Ser Ser Glu Phe Gly Arg Arg Val Trp Thr Ser
      20      25      30
Ala Pro Pro Pro Gln Arg Pro Phe Arg Val Cys Asp His Lys Arg
      35      40      45
Thr Ile Arg Lys Gly Leu Thr Ala Ala Thr Arg Gln Glu Leu Leu
      50      55      60
Ala Lys Ala Leu Glu Thr Leu Leu Leu Asn Gly Val Leu Thr Leu
      65      70      75
Val Leu Glu Glu Asp Gly Thr Ala Val Asp Ser Glu Asp Phe Phe
      80      85      90
Gln Leu Leu Glu Asp Asp Thr Cys Leu Met Val Leu Gln Ser Gly
      95      100      105
Gln Ser Trp Ser Pro Thr Arg Ser Gly Val Leu Ser Tyr Gly Leu
      110      115      120
Gly Arg Glu Arg Pro Lys His Ser Lys Asp Ile Ala Arg Phe Thr
      125      130      135
Phe Asp Val Tyr Lys Gln Asn Pro Arg Asp Leu Phe Gly Ser Leu
      140      145      150
Asn Val Lys Ala Thr Phe Tyr Gly Leu Tyr Ser Met Ser Cys Asp

```

	155		160		165
Phe Gln Gly Leu Gly	Pro Lys Lys Val	Leu Arg Glu Leu Leu Arg			
	170		175		180
Trp Thr Ser Thr Leu	Leu Gln Gly Leu Gly	His Met Leu Leu Gly			
	185		190		195
Ile Ser Ser Thr Leu	Arg His Ala Val Glu	Gly Ala Glu Gln Trp			
	200		205		210
Gln Gln Lys Gly Arg	Leu His Ser Tyr				
	215				

&lt;210&gt; 16

&lt;211&gt; 439

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 001273CD1

&lt;400&gt; 16

Met Ala Ala Ala Arg	Cys Trp Arg Pro	Leu Leu Arg Gly Pro Arg	
1	5	10	15
Leu Ser Leu His Thr	Ala Ala Asn Ala Ala	Ala Thr Ala Thr Glu	
	20	25	30
Thr Thr Cys Gln Asp	Val Ala Ala Thr Pro	Val Ala Arg Tyr Pro	
	35	40	45
Pro Ile Val Ala Ser	Met Thr Ala Asp Ser	Lys Ala Ala Arg Leu	
	50	55	60
Arg Arg Ile Glu Arg	Trp Gln Ala Thr Val	His Ala Ala Glu Ser	
	65	70	75
Val Asp Glu Lys Leu	Arg Ile Leu Thr Lys	Met Gln Phe Met Lys	
	80	85	90
Tyr Met Val Tyr Pro	Gln Thr Phe Ala Leu	Asn Ala Asp Arg Trp	
	95	100	105
Tyr Gln Tyr Phe Thr	Lys Thr Val Phe Leu	Ser Gly Leu Pro Pro	
	110	115	120
Arg Pro Ser Glu Pro	Glu Pro Glu Pro Glu	Pro Glu Pro Glu Pro	
	125	130	135
Ala Leu Asp Leu Ala	Ala Leu Arg Ala Val	Ala Cys Asp Cys Leu	
	140	145	150
Leu Gln Glu His Phe	Tyr Leu Arg Arg Arg	Arg Arg Arg Val His Arg	
	155	160	165
Tyr Glu Glu Ser Glu	Val Ile Ser Leu Pro	Phe Leu Asp Gln Leu	
	170	175	180
Val Ser Thr Leu Val	Gly Leu Leu Ser Pro	His Asn Pro Ala Leu	
	185	190	195
Ala Ala Ala Ala Leu	Asp Tyr Arg Cys Pro	Val His Phe Tyr Trp	
	200	205	210
Val Arg Gly Glu Glu	Ile Ile Pro Arg Gly	His Arg Arg Gly Arg	
	215	220	225
Ile Asp Asp Leu Arg	Tyr Gln Ile Asp Asp	Lys Pro Asn Asn Gln	
	230	235	240
Ile Arg Ile Ser Lys	Gln Leu Ala Glu Phe	Val Pro Leu Asp Tyr	
	245	250	255
Ser Val Pro Ile Glu	Ile Pro Thr Ile Lys	Cys Lys Pro Asp Lys	

	260		265		270
Leu Pro Leu Phe	Lys Arg Gln Tyr Glu	Asn His Ile Phe Val	Gly		
	275		280		285
Ser Lys Thr Ala	Asp Pro Cys Cys Tyr	Gly His Thr Gln Phe	His		
	290		295		300
Leu Leu Pro Asp	Lys Leu Arg Arg Glu	Arg Leu Leu Arg Gln	Asn		
	305		310		315
Cys Ala Asp Gln	Ile Glu Val Val Phe	Arg Ala Asn Ala Ile	Ala		
	320		325		330
Ser Leu Phe Ala	Trp Thr Gly Ala Gln	Ala Met Tyr Gln Gly	Phe		
	335		340		345
Trp Ser Glu Ala	Asp Val Thr Arg Pro	Phe Val Ser Gln Ala	Val		
	350		355		360
Ile Thr Asp Gly	Lys Tyr Phe Ser Phe	Phe Cys Tyr Gln Leu	Asn		
	365		370		375
Thr Leu Ala Leu	Thr Thr Gln Ala Asp	Gln Asn Asn Pro Arg	Lys		
	380		385		390
Asn Ile Cys Trp	Gly Thr Gln Ser Lys	Pro Leu Tyr Glu Thr	Ile		
	395		400		405
Glu Asp Asn Asp	Val Lys Gly Phe Asn	Asp Asp Val Leu Leu	Gln		
	410		415		420
Ile Val His Phe	Leu Leu Asn Arg Pro	Lys Glu Glu Lys Ser	Gln		
	425		430		435
Leu Leu Glu Asn					

&lt;210&gt; 17

&lt;211&gt; 526

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 411831CD1

&lt;400&gt; 17

Met Ala Ser Gly Pro	His Ser Thr Ala Thr	Ala Ala Ala Ala Ala	
1	5	10	15
Ser Ser Ala Ala Pro	Ser Ala Gly Gly Ser	Ser Ser Ser Gly Thr	Thr
	20	25	30
Thr Thr Thr Thr Thr	Thr Thr Gly Gly Ile	Leu Ile Gly Asp Arg	
	35	40	45
Leu Tyr Ser Glu Val	Ser Leu Thr Ile Asp	His Ser Leu Ile Pro	
	50	55	60
Glu Glu Arg Leu Ser	Pro Thr Pro Ser Met	Gln Asp Gly Leu Asp	
	65	70	75
Leu Pro Ser Glu Thr	Asp Leu Arg Ile Leu	Gly Cys Glu Leu Ile	
	80	85	90
Gln Ala Ala Gly Ile	Leu Leu Arg Leu Pro	Gln Val Ala Met Ala	
	95	100	105
Thr Gly Gln Val Leu	Phe His Arg Phe Phe	Tyr Ser Lys Ser Phe	
	110	115	120
Val Lys His Ser Phe	Glu Ile Val Ala Met	Ala Cys Ile Asn Leu	
	125	130	135
Ala Ser Lys Ile Glu	Glu Ala Pro Arg Arg	Ile Arg Asp Val Ile	

Asn Val Phe His	140	145	150
His Leu Arg Gln Leu Arg Gly Lys Arg Thr Pro	155	160	165
Ser Pro Leu Ile Leu Asp Gln Asn Tyr Ile Asn Thr Lys Asn Gln	170	175	180
Val Ile Lys Ala Glu Arg Arg Val Leu Lys Glu Leu Gly Phe Cys	185	190	195
Val His Val Lys His Pro His Lys Ile Ile Val Met Tyr Leu Gln	200	205	210
Val Leu Glu Cys Glu Arg Asn Gln Thr Leu Val Gln Thr Ala Trp	215	220	225
Asn Tyr Met Asn Asp Ser Leu Arg Thr Asn Val Phe Val Arg Phe	230	235	240
Gln Pro Glu Thr Ile Ala Cys Ala Cys Ile Tyr Leu Ala Ala Arg	245	250	255
Ala Leu Gln Ile Pro Leu Pro Thr Arg Pro His Trp Phe Leu Leu	260	265	270
Phe Gly Thr Thr Glu Glu Glu Ile Gln Glu Ile Cys Ile Glu Thr	275	280	285
Leu Arg Leu Tyr Thr Arg Lys Lys Pro Asn Tyr Glu Leu Leu Glu	290	295	300
Lys Glu Val Glu Lys Arg Lys Val Ala Leu Gln Glu Ala Lys Leu	305	310	315
Lys Ala Lys Gly Leu Asn Pro Asp Gly Thr Pro Ala Leu Ser Thr	320	325	330
Leu Gly Gly Phe Ser Pro Ala Ser Lys Pro Ser Ser Pro Arg Glu	335	340	345
Val Lys Ala Glu Glu Lys Ser Pro Ile Ser Ile Asn Val Lys Thr	350	355	360
Val Lys Lys Glu Pro Glu Asp Arg Gln Gln Ala Ser Lys Ser Pro	365	370	375
Tyr Asn Gly Val Arg Lys Asp Ser Lys Arg Ser Arg Asn Ser Arg	380	385	390
Ser Ala Ser Arg Ser Arg Ser Arg Thr Arg Ser Arg Ser Arg Ser	395	400	405
His Thr Pro Arg Arg His Tyr Asn Asn Arg Arg Ser Arg Ser Gly	410	415	420
Thr Tyr Ser Ser Arg Ser Arg Ser Arg Ser Arg Ser His Ser Glu	425	430	435
Ser Pro Arg Arg His His Asn His Gly Ser Pro His Leu Lys Ala	440	445	450
Lys His Thr Arg Asp Asp Leu Lys Ser Ser Asn Arg His Gly His	455	460	465
Lys Arg Lys Lys Ser Arg Ser Arg Ser Gln Ser Lys Ser Arg Asp	470	475	480
His Ser Asp Ala Ala Lys Lys His Arg His Glu Arg Gly His His	485	490	495
Arg Asp Arg Arg Glu Arg Ser Arg Ser Phe Glu Arg Ser His Lys	500	505	510
Ser Lys His His Gly Gly Ser Arg Ser Gly His Gly Arg His Arg	515	520	525

Arg

&lt;210&gt; 18

```
<220>
<221> misc_feature
<223> Incyte ID No: 1520835CD1
```

Met	Gly	Pro	Lys	Asp	Ser	Ala	Lys	Cys	Leu	His	Arg	Gly	Pro	Gln
1				5					10					15
Pro	Ser	His	Trp	Ala	Ala	Gly	Asp	Gly	Pro	Thr	Gln	Glu	Arg	Cys
				20					25					30
Gly	Pro	Arg	Ser	Leu	Gly	Ser	Pro	Val	Leu	Gly	Leu	Asp	Thr	Cys
				35					40					45
Arg	Ala	Trp	Asp	His	Val	Asp	Gly	Gln	Ile	Leu	Gly	Gln	Leu	Arg
				50					55					60
Pro	Leu	Thr	Glu	Glu	Glu	Glu	Glu	Glu	Gly	Ala	Gly	Ala	Thr	Leu
				65					70					75
Ser	Arg	Gly	Pro	Ala	Phe	Pro	Gly	Met	Gly	Ser	Glu	Glu	Leu	Arg
				80					85					90
Leu	Ala	Ser	Phe	Tyr	Asp	Trp	Pro	Leu	Thr	Ala	Glu	Val	Pro	Pro
				95					100					105
Glu	Leu	Leu	Ala	Ala	Ala	Gly	Phe	Phe	His	Thr	Gly	His	Gln	Asp
				110					115					120
Lys	Val	Arg	Cys	Phe	Phe	Cys	Tyr	Gly	Gly	Leu	Gln	Ser	Trp	Lys
				125					130					135
Arg	Gly	Asp	Asp	Pro	Trp	Thr	Glu	His	Ala	Lys	Trp	Phe	Pro	Ser
				140					145					150
Cys	Gln	Phe	Leu	Leu	Arg	Ser	Lys	Gly	Arg	Asp	Phe	Val	His	Ser
				155					160					165
Val	Gln	Glu	Thr	His	Ser	Gln	Leu	Leu	Gly	Ser	Trp	Asp	Pro	Trp
				170					175					180
Glu	Glu	Pro	Glu	Asp	Ala	Ala	Pro	Val	Ala	Pro	Ser	Val	Pro	Ala
				185					190					195
Ser	Gly	Tyr	Pro	Glu	Leu	Pro	Thr	Pro	Arg	Arg	Glu	Val	Gln	Ser
				200					205					210
Glu	Ser	Ala	Gln	Glu	Pro	Gly	Gly	Val	Ser	Pro	Ala	Glu	Ala	Gln
				215					220					225
Arg	Ala	Trp	Trp	Val	Leu	Glu	Pro	Pro	Gly	Ala	Arg	Asp	Val	Glu
				230					235					240
Ala	Gln	Leu	Arg	Arg	Leu	Gln	Glu	Glu	Arg	Thr	Cys	Lys	Val	Cys
				245					250					255
Leu	Asp	Arg	Ala	Val	Ser	Ile	Val	Phe	Val	Pro	Cys	Gly	His	Leu
				260					265					270
Val	Cys	Ala	Glu	Cys	Ala	Pro	Gly	Leu	Gln	Leu	Cys	Pro	Ile	Cys
				275					280					285
Arg	Ala	Pro	Val	Arg	Ser	Arg	Val	Arg	Thr	Phe	Leu	Ser		
				290					295					

```
<210> 19
<211> 249
<212> PRT
<213> Homo sapiens
```



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1902803CD1

&lt;400&gt; 19

```

Met Ala Ala Gly Met Tyr Leu Glu His Tyr Leu Asp Ser Ile Glu
 1          5          10          15
Asn Leu Pro Phe Glu Leu Gln Arg Asn Phe Gln Leu Met Arg Asp
          20          25          30
Leu Asp Gln Arg Thr Glu Asp Leu Lys Ala Glu Ile Asp Lys Leu
          35          40          45
Ala Thr Glu Tyr Met Ser Ser Ala Arg Ser Leu Ser Ser Glu Glu
          50          55          60
Lys Leu Ala Leu Leu Lys Gln Ile Gln Glu Ala Tyr Gly Lys Cys
          65          70          75
Lys Glu Phe Gly Asp Asp Lys Val Gln Leu Ala Met Gln Thr Tyr
          80          85          90
Glu Met Val Asp Lys His Ile Arg Arg Leu Asp Thr Asp Leu Ala
          95          100          105
Arg Phe Glu Ala Asp Leu Lys Glu Lys Gln Ile Glu Ser Ser Asp
          110          115          120
Tyr Asp Ser Ser Ser Ser Lys Gly Lys Lys Lys Gly Arg Thr Gln
          125          130          135
Lys Glu Lys Lys Ala Ala Arg Ala Arg Ser Lys Gly Lys Asn Ser
          140          145          150
Asp Glu Glu Ala Pro Lys Thr Ala Gln Lys Lys Leu Lys Leu Val
          155          160          165
Arg Thr Ser Pro Glu Tyr Gly Met Pro Ser Val Thr Phe Gly Ser
          170          175          180
Val His Pro Ser Asp Val Leu Asp Met Pro Val Asp Pro Asn Glu
          185          190          195
Pro Thr Tyr Cys Leu Cys His Gln Val Ser Tyr Gly Glu Met Ile
          200          205          210
Gly Cys Asp Asn Pro Asp Cys Ser Ile Glu Trp Phe His Phe Ala
          215          220          225
Cys Val Gly Leu Thr Thr Lys Pro Arg Gly Lys Trp Phe Cys Pro
          230          235          240
Arg Cys Ser Gln Glu Arg Lys Lys Lys
          245

```

&lt;210&gt; 20

&lt;211&gt; 1748

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1342011CB1

&lt;400&gt; 20

```

cgggtcgacc cagcggtccg gggggacaag ccagaggctg gaggagcagc atcccttcca 60
ggctgcacac ttgtcagtgc cgggttcttg ggagaaccgc acgggaagga gaggtcgctg 120
gcggtcatcgt ttgtgtctcc ccagagacag acctgggccc ttccctcttg gactcccaat 180
ctggacgggg ttcttggtt gctgtggggc atgttgaggc cggaggcttg gcttgtgggg 240

```

```

ctgcacggcc ctgcccagga gaactcagca ctgcctggac ggtgaggctc agcttctgag 300
ctgagggctc tatcaggcct ggaagtggac cctggggagg ggtggggcag ggtagttctg 360
ataagtecta ggactgttcg cttccgggtt ctgagccctg gcgtcaggga ggaagggcat 420
gtccagaaca atggccagaa ccaggcccg ccagctcggg cgggtgacgg gggcggttg 480
ctggggcagc gctgcccgtg gcaggggccg agccctgcgg ggccgtgagc cggccctgcc 540
ttctgcttcc tttccagatg tagccgctg tcccgggagc ctggactgtg ccctgaagag 600
gcgggcaagg tgtcctcctg gtgcacatgc ctgtgggccc tgccttcagc cctccagga 660
ggaccagcaa gggctctgtg tgcccaggat gcgcggcct ccaggcggg gcccggccca 720
gcccagactg gaagatgaga ttgacttcct ggcccaggag cttgcccga aggagtctg 780
aactcaact ccgcccctac ccaaggaccg acagcgctc ccggagcctg ccacctggg 840
cttctcgga cgggggcagg ggttgagct gggcctcccc tccactccag gaacccccac 900
gcccacgccc cacacctccc tgggtcccc tgtgtcatcc gaccgggtgc acatgtgcc 960
cctggagccc cggggagggc aaggcgacgg cctcgccctt gtgctgatcc tggcgttctg 1020
tgtggcgggt gcagccgccc tctccgtagc ctccctctgc tgggtgcaggc tgcagcgtga 1080
gatccgctg actcagaagg ccgactacgc cactgcgaag gcccctggct cacctgcagc 1140
tccccgata tgcctgggg accagcggt ggacagagc gcggagatgt accactacca 1200
gcaccaacgg caacagatgc tgtgcctgga gcgcataaa gagccacca aggagctgga 1260
cacggcctcc tcggtgagg agaattgagga cggagacttc acggtgtacg agtgccggg 1320
cctggccccc accggggaaa tggaggtgcg caaccctctg ttcgaccacg ccgactgtc 1380
cgcccccctg ccggccccc gctcaccgcc tgcactgcca tgacctggag gcagacagac 1440
gcccactgc tccccgacct cgaggccccc ggggaggggc agggcctgga gcttcccact 1500
aaaaacatgt tttgatgctg tgtgcttttg gctgggccc gggctccagg ccctgggacc 1560
ccttgccagg gagaccccc aacctttgtg ccaggacacc tctggtccc ctgcacctct 1620
cctgttcggg ttagaccccc aaactggagg gggcatggag aaccgtagag cgcaggaacg 1680
ggtgggtaat tctagagaca aaagccaatt aaagtccatt tcagacctgc aaaaaaaaaa 1740
aaaaaagg                                     1748

```

&lt;210&gt; 21

&lt;211&gt; 1016

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1880041CB1

&lt;400&gt; 21

```

ccgcagtctc tgtgcgttga agccggagac cgcggcgccc tcagcgagga ccctccgccc 60
cggagccgcc ggccggagcc gcagcctctg ccgcagcgcc cccgccacct gtcccctccc 120
cctccgcctc cgcgggagcc gcctcgtgca ctctggggta tggcgtcaa tgtgtactcc 180
acatctgtga ccagtgaata tctgagtcgc catgatatgc ttgcatgggt caacgactcc 240
ctgcacctca actataccaa gatagaacag ctttgttcag gggcagccta ctgccagttc 300
atggacatgc ttttccccg ctgtgtgac ttgaggaaag tgaagtcca ggccaaacta 360
gagcatgaat acatccacaa cttcaagggt ctgcaagcag ctttcaagaa gatgggtgtt 420
gacaaaatca ttcctgtaga gaaattagtg aaaggaaaa tccaagataa ttttgagttt 480
attcagtggg ttaagaaatt ctttgacgca aactatgatg gaaaggatta caacctctg 540
ctggcgcgcc agggccagga cgtagcgcca cctcctaacc caggtgatca gatcttcaac 600
aaatccaaga aactcattgg cacagcagtt ccacagagga cgtccccccac agggccaaaa 660
aacatgcaga cctctggccg gctgagcaat gtggccccc cctgcattct ccggaagaat 720
cctccatcag ccgaaatgg cggccatgag actgatgccc aaattcttga actcaaccaa 780
cagctggtgg acttgaagct gacagtggat gggctggaga aggaacgtga cttctacttc 840
agcaaacttc gtgacatcga gctcatctgc caggagcatg aaagtgaata cagccctgtt 900
atctcaggca tcattggcat cctctatgcc acagaggaag gattcgcacc ccctgaggac 960
gatgagattg aagagcatca acaagaagac caggacgagt actgagggcg gccgca 1016

```

<210> 22  
 <211> 1145  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3201881CB1

<400> 22  
 tgtgccccaga acgcgggttag gaagtgtgtg catacgtctg aaccctaaat ggttctcagt 60  
 tctgtaaact tctcctccca ctgggtggag tagggccttt aagagcagct ggaatgcagt 120  
 tcccctgatc agcgtaccag ttgttgctct tctgaacctc tgccagtcct ggagactggg 180  
 gccctgagct ccaaccagcg ggcctcatcc tacacctca ccaccgcaac ttctcaccgg 240  
 agcaagaagc agtcccaga gagaaagaac gttcccacct gcctagccat gggagaggac 300  
 gctgcacagg ccgaaaagtt ccagcacctt ggggtctgaca tgcggcagga aaagccctcg 360  
 agccccagcc cgtatgccttc ctccacacca agccccagcc tgaacctagg gaacacagag 420  
 gaggccatcc gggacaactc acaggtgaac gcagtcacgg tgctcacgct cctggacaag 480  
 ctgggtgaaca tgctagacgc tgtgcaggag aaccagcaca agatggagca gcgacagatc 540  
 agttttggagg gctccgtgaa gggcatccag aatgacctca ccaagctctc caagtaccag 600  
 gcctccacca gcaaacagggt gagcaagctg ctggagaagt cccgcaagggt cagcgcccac 660  
 acgcgcgcgg tcaaaagagcg catggatagg cagtgcgcac aggtgaagcg gctggagaac 720  
 aaccacgccc agtcctccg acgcaaccat ttcaaagtgc tcatcttcca ggaggaaaat 780  
 gagatccctg ccagcgtggt tgtgaaacag cccgtttccg gtgccgtgga agggaaggag 840  
 gagcttccgg atgaaaacaa atccctggag gaaacctgc acaccgtgga cctctcctca 900  
 gatgatgatt tgccccacga tgaggaggcc ctggaagaca gtgccgagga aaaggttgga 960  
 agaagtaggg gcagagaaat taaaagatcc cggccgtgaa ggaagttgga tagcctcaaa 1020  
 gaaagctttt tttttgccag gaactttggg gaaaaagggt gaacaagtct gggggacaaa 1080  
 gttcctatct tgtaggaga aggggagagg agtttttaga atcttcttca agtcaaactc 1140  
 accag 1145

<210> 23  
 <211> 3084  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 939000CB1

<400> 23  
 cggctcgagg cgggtgacgc gacgacggcg acactttgct acggagtgca ttcggacgtc 60  
 gaagcctaga gtctctgcgt ctttccctct tccgctgcct cattcctttc cttectagcc 120  
 ttggtcgtcg ccgccaccat gaacaagaag aagaaaaccgt tcctagggat gcccgcgccc 180  
 ctcggtacg tgcgggggct gggccggggc gccactggct tcaccacgcg gtcagacatt 240  
 gggcccgccc gtgatgcaaa tgacctgtg gatgatcgcc atgcaccccc aggcaagaga 300  
 accgttgggg accagatgaa gaaaaatcag gctgctgacg atgacgacga ggatctaaat 360  
 gacaccaatt acgatgagtt taatggctat gctgggagcc tcttctcaag tggacctac 420  
 gagaaagatg atgaggaagc agatgctatc tatgcagccc tggataaaag gatggatgaa 480  
 agaagaaaag aaagacggga gcaaaggag aaagaagaaa tagagaaata tcgtatggaa 540  
 cgccccaaaa tccaacagca gttctcagac ctcaagagga agttggcaga agtcacagaa 600  
 gaagagtggc tgagcatccc cgaggttggc gatgccagaa ataaacgtca gcggaaccca 660  
 cgctatgaga agctgacccc tgttcctgac agtttctttg ccaaacttt acagaccgga 720  
 gagaaccata cctcagtgga tccccgacaa actcaatttg gaggtcttaa cacacctat 780

```

ccagggtggac taaacactcc ataccaggt ggaatgacgc caggactgat gacacctggc 840
acagggtgagc tggacatgag gaagattggc caagcgagga acactctgat ggacatgagg 900
ctgagccagg tgtctgactc cgtgagtggg cagaccgtcg ttgaccccaa aggctacctg 960
acggattttaa attccatgat cccgacacac ggaggagaca tcaatgatat caagaaggcg 1020
cgactgctcc tcaagtctgt tcgggagacg aacctctatc acccgccagc ctggattgca 1080
tcagcccgcg tggaaagaag cactgggaag ctacaagtag ctcggaacct tatcatgaag 1140
gggacggaga tgtgccccaa gagtgaagat gtctggctgg aagcagccag gttgcagcct 1200
gggacacacg ccaaggccgt ggtagcccaa gctgtccgtc atctccaca gtctgtcagg 1260
atttacatca gagccgcaga gctggaaacg gacattcgtg caaagaagcg ggttcttcg 1320
aaagccctcg agcatgttcc aaactcgggt cgcttgtgga aagcagccgt tgagctggaa 1380
gaacctgaag atgctagaat catgctgagc cgagctgtgg agtctgtccc caccagcgtg 1440
gagctctggc ttgctctggc aaggctggag acctatgaaa atgcccga ggtcttgaac 1500
aaggcgcgga agaacattcc tacagaccga catatctgga tcacggctgc taagctggag 1560
gaagccaatg ggaacacgca gatggtggag aagatcatcg accgagccat cacctcgtg 1620
cgggccaacg gtgtggagat caaccgtgag cagtggatcc aggatgccga ggaatgtgac 1680
agggtcggga gtgtggccac ctgccaggcc gtcattcgtg ccgtgattgg gattgggatt 1740
gaggaggaag atcggaagca tacctggatg gaggatgctg acagtttgtg agcccacaat 1800
gccctggagt gtgcacgagc catctacgcc tacgccctgc aggtgttccc cagcaagaag 1860
agtgtgtggc tgcgcgccgc gtacttcgag aagaacctg gcactcggga gtccctggaa 1920
gacctcctgc agtggtctgt gggccactgc cccaaagcag aggtgctgtg gctcatggc 1980
gccaaagcca agtggctggc aggggatgtg cctgcagcaa ggagcatcct ggccctggcc 2040
ttccaggcca accccaacag tgaggagatc tggctggcag ccgtgaagct ggagtcagag 2100
aatgatgagt acgagcgggc cggagggtg ctggccaagg cgcggagcag tgccccacc 2160
gcccgggtgt tcatgaagtc tgtgaagctg gagtgggtgc aagacaacat cagggcagcc 2220
caagatctgt gcgaggagc cctgcggcac tatgaggact tccccaagct gtggatgatg 2280
aaggggcaga tcgaggagca gaaggagatg atggagaagg cgcgggaagc ctataaccag 2340
gggttgaaga agtgtcccca ctccacaccc ctgtggcttt tgctctctcg gctggaggag 2400
aagattgggc agcttactcg agcacgggcc attttggaaa agtctcgtct gaagaaccca 2460
aagaaccctg ggctgtggtt ggagtcctg cggtggagt accgtgcggg gctgaagaac 2520
atcgcaaata cactcatggc caaggcgtg caggagtgc ccaactccg taccctgtgg 2580
tctgaggcca tcttctcga ggaaggccc cagaggagga ccaagagcgt ggatgccctg 2640
aagaagtgtg agcatgacct ccattgtctc ctggccgtgg ccaagctgtt ttggagtcag 2700
cggaaagatca ccaaggccag ggagtggttc caccgactg tgaagattga ctggacctg 2760
ggggatgcct gggccttctt ctacaagttt gagctgcagc atggcactga ggagcagcag 2820
gaggaggtga ggaagcgtg tgagagtgc gagcctcggc atggggagct gtggtgcgcc 2880
gtgtccaagg acatcgccaa ctggcagaag aagatcgggg acatccttag gctggtggcc 2940
ggccgcatca agaacacct ctgattgagc ggttgccatg gccggtctcc gtggggcagg 3000
gttgggcgcg atgtggaagg gctctgagct gtgtcctcct tcattaaaag tttttatgtc 3060
tcgtgtcaga aaaaaaaaaa aaaa 3084

```

&lt;210&gt; 24

&lt;211&gt; 3315

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2125677CB1

&lt;400&gt; 24

```

acttattccc acttggaaact ggatggtcag tattatggat tctactgaag ctcaattacg 60
ttatggttct gcattagcat ctgctgggtga tcctggacat ccaaagcatc ctcttcacga 120
ttctcagaat tcagcgagaa gagagaggat gactgcgcga gaagaagcta gcttacgaac 180
acttgaaggc agacgacgtg ccaccttgct tagcgcctcg caaggaatga tgtctgcacg 240

```

```

aggagacttc ctaaattatg ctctgtctct aatgcggtct cataatgatg agcattctga 300
tgttcttcca gttttggatg tttgtctatt gaagcatgtg gcatatgttt ttcaagcact 360
tataactggg attaaggcaa tgaatcagca gacaacattg gatacacctc aactagaacg 420
caaaaggacg cgagaactct tggaaactggg tattgataat gaagattcag aacatgaaaa 480
tgatgatgac accaaactaaa gtgctacttt gaatgataag gatgatgact ctcttctctg 540
agaaactggc caaaaccatc catttttccg acgttcagac tccatgacat tccctgggtg 600
tataccccca aatccatttg aagtgcctct ggctgaagcc atcccccttg ctgatcagcc 660
acatctgttg cagccaaatg ctagaaagga ggatcttttt ggccgtccaa gtcagggtct 720
ttattcttca tctgccagta gtgggaaatg tttaatggag gttacagtgg atagaaactg 780
cctagagggt cttccaacaa aaatgtctta tgctgccaat ctgaaaaatg taatgaacat 840
gcaaaaccgg caaaaaaaag aaggggaaga acagcccgtg ctgccagaag aaactgagag 900
ttcaaaacca gggccatctg ctcatgatct tgctgcacaa ttaaaaagta gcttactagc 960
agaaatagga cttactgaaa gtgaagggcc acctctcaca tctttcaggc cacagtgtag 1020
ctttatggga atgggtatct cccatgatat gctgctagga cgttggcgcc tttctttaga 1080
actgttcggc aggggtattca tgggaagatgt tggagcagaa cctggatcaa tccaaactga 1140
attgggtggt tttgaggtaa aagaatcaaa attccgcaga gaaatggaaa aactgagaaa 1200
ccagcagtc agagatttgt cactagagggt aaaggttgat cgggatcgag atcttctcat 1260
tcagcagact atgaggcagc ttaacaatca ctttggtcga agatgtgcta ctacaccaat 1320
ggctgtacac agagtaaaaag tcacatttaa ggatgagcca ggagagggca gtggtgtagc 1380
acgaagtctt tatcacagca ttgcacaagc atttttatca aatgaaaaat tgccaaatct 1440
agagtgtatc caaaatgcc acaaaggcac ccacacaagt ttaatgcaga gattaaggaa 1500
ccgaggagag agagaccggg aaagggagag agaaagggaa atgaggagga gtagtgggtt 1560
gagagcaggt tctcgagggt accgggatag agactttaga agacagcttt ccatcgacac 1620
taggcccttt agaccagcct ctgaagggaa tctagcgat gatcctgagc ctttgccagc 1680
acatcggcag gcacttgag agaggcttta tctctgtgta caagcaatgc aaccagcatt 1740
tgcaagtaaa atcactggca tgttgttgga attatcccca gctcagctgc ttctccttct 1800
agcaagtga gattctctga gagcaagagt ggatgaggcc atggaaactca ttattgcaca 1860
tggacgggaa aatggagctg atagtatcct ggatcttgga ttagtagact cctcagaaaa 1920
ggtacagcag gaaaaccgaa agcgccatgg ctctagtcga agtgtagtag atatggattt 1980
agatgataca gatgatggtg atgacaatgc cctttgttt taccaacctg ggaaaagagg 2040
attttatact ccaaggcctg gcaagaacac agaagcaagg ttgaattggt tcagaaacat 2100
tggcaggatt ctgggactat gtctgttaca gaatgaacta tgtcctatca cattgaatag 2160
acatgtaat aaagtattgc ttggtagaaa agtcaattgg catgattttg ctttttttga 2220
tctgtaatg tatgagagtt tgcggcaact aatcctcgcg tctcagagtt cagatgctga 2280
tgctgttttc tcagcaatgg atttggcatt tgcaattgac ctgtgtaaag aagaagggtg 2340
aggacagggt gaactcattc ctaatggtgt aaatatacca gtcactccac agaattgata 2400
tgagtatgtg cggaaatacg cagaacacag aatgttggtg gttgcagaac agcccttaca 2460
tgcaatgagg aaaggtctac tagatgtgct tccaaaaaat tcattagaag atttaacggc 2520
agaagatttt aggccttttg taaatggctg cggtgaaagc aatgtgcaaa tgctgatcag 2580
ttttacctct ttcaatgatg aatcaggaga aaatgctgag aagcttctgc agttcaagcg 2640
ttggttctgg tcaatagtag agaagatgag catgacagaa cgacaagatc ttgtttactt 2700
ttggacatca agcccatcac tgccagccag tgaagaagga ttccagccta tgccctcaat 2760
cacaataaga ccaccagatg accaaccatct tctactgca aatacttgca tttctcgact 2820
ttacgtccca ctctattcct ctaaacagat tctcaaacag aaattgttac tcgccattaa 2880
gaccaagaat tttggttttg ttagagtagt aaaaagtgtg tattgctgtg taatattact 2940
agcaaatttt gtgatttttt ttccatttgt ctataaaaagt ttatggaagt taatgctgtc 3000
ataccccctt ggtggtacct taaagagata aaatgcagac attccttgct gagtttatag 3060
cttaaaggcc taaggagcac tagcaacatt tggctatatt ggtttgctag tcaccaactt 3120
ctgggtctaa cccagccaa agatgacagc agaacaacat aatttacact gtgatttatc 3180
tttttgctga gggggaaaaa atgtaaatgt tctgaaaatt cactgctgcc tttgtggaaa 3240
ctgtttcagc aaaggttctt gtatagaggg aatagggaaat ttcaaaataa aaaattaagt 3300
atgaaaaaaa aaaaa 3315

```

<211> 1677  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2603810CB1

<400> 25  
 tgggaggggg cggaattcc cgactctagg ccggaagcgc ggggagacca tgtagtgaga 60  
 ccctcgcgag gtctgagagt cactggagct accagaagca tcatggggcc ctggggagag 120  
 ccagagctcc tgggtgtggcg ccccagggcg gttagcttcag agcctccagt gcctgtgggg 180  
 ctggaggtga agttgggggc cctgggtgctg ctgctgggtgc tcacctcct ctgcagcctg 240  
 gtgccatct gtgtgctgcg ccggccagga gctaaccatg aaggtcagc tccccgccag 300  
 aaagccctga gcctagtaag ctgtttcgcg gggggcgtct ttttgccac ttgtctcctg 360  
 gacctgctgc ctgactacct ggctgccata gatgaggccc tggcagcctt gcacgtgacg 420  
 ctccagttcc cactgcaaga gttcctcctg gccatgggct tcttcctggt cctggtgatg 480  
 gacgagatca cactggctta caaggagcag tcaggggcgt cacctctgga ggaaacaagg 540  
 gctctgctgg gaacagtga tgggtgggccc cagcattggc atgatgggcc aggggtccca 600  
 caggcgagtg gagccccagc aacccccctca gccttgctg cctgtgtact ggtgttctcc 660  
 ctggccctcc actccgtgtt cgaggggctg gcgttagggc tgcagcgaga ccgggctcag 720  
 gccatggagc tgtgcctggc tttgctgctc cacaagggca tectggctgt cagcctgtcc 780  
 ctgctggctgt tgcagagcca ccttagggca cagggtggtg ctggctgtgg gatectcttc 840  
 tcatgcatga cacctctagg catcgggctg ggtgcagctc tggcagagtc ggcaggacct 900  
 ctgcaccagc tggcccagtc tgtgctagag ggcatggcag ctggcacctt tctctatata 960  
 acctttctgg aaatcctgcc ccaggagctg gccagttctg agcaaaggat cctcaaggctc 1020  
 attctgctcc tagcaggett tgccctgctc actggcctgc tcttcacca aatctagggg 1080  
 gcttcaagag aggggcaggg gagattgatg atcaggtgcc cctgttctcc ctccctccc 1140  
 ccagttgtgg ggaataggaa ggaaagggga agggaaatac tgaggaccaa aaagttctct 1200  
 gggagctaaa gatagagcct ttggggctat ctgactaatg agagggaagt gggcagacaa 1260  
 gaggtgggcc ccagtcccaa ggaacaagag atggtcaagt cgctagagac atatcagggg 1320  
 acattaggat tggggaagac acttgactgc tagaatcaga ggttggaacac tatacataag 1380  
 gacaggttca ctaggaggc tggaggtggg taccagctg ctgtggaacg ggtatggaga 1440  
 ggtcataaac ctatagtcag tgtcctggtg tccctagccc atttcagcac cctgccactt 1500  
 ggagtggacc cctcctactc ttcttagcgc ctacctcat acctatctcc ctccctccat 1560  
 ctccagggg actggcgcca aatgggtctc cctgccaat tttggaatc tctctggcct 1620  
 ctccagtcct gcttactccc ctatttttaa agtgccaaac aatccccctc ctcttttc 1677

<210> 26  
 <211> 997  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2715761CB1

<400> 26  
 gcgacccttg ttcaacgccg ttggcgaaca gctgctggag gtgccgagaa tctgagtttc 60  
 ggcaagcagc caggtctgga aactaatatt taaaaaatga ctacaccaa caagacacct 120  
 cctgggtgctg accccaagca gttggaaagg actggaacag tacgggaaat tgggtcacia 180  
 gctgtttggg cactctcatc ttgcaaacca ggatttggag tggatcagtt acgagatgac 240  
 aatctagaaa cttattggca atcagatggg tcccagcctc atttagtgaa catccaattc 300  
 agaagaaaaa caacagtga gacattatgt atttatgcag actacaaatc tgatgaaagc 360

```

tatactccaa gcaagatctc agtcagagta ggaaataaatt ttcacaacct tcaagaaatt 420
cggcaacttg agttgggtgga accaagtggc tggattcatg ttcccttaac tgacaatcat 480
aagaagccaa ctctgtacatt catgatacag attgctgttc tagccaatca ccagaatgga 540
agagacaccc atatgagaca aattaaaaata tacacaccag tagaagagag ctccattggt 600
aaatttcccta gatgtacaac tatagatttc atgatgtatc gttcaataag gtgactttaa 660
aatgagacga aaatcattaa acgtatcttt gttttatcct gtattttaa aatatatcat 720
gtacctttat tgaacaaggc atccgtttata tctaattttg tatatgttta aaaatatttt 780
attgtaactt tgacaaaata atttggggtc atattatctt tattttcttt aacatgtaat 840
aaagctcaca tattttacat tactaaaaat ggatttgaag ccaatcattt tattttccct 900
tgtatcaaaa gaaaagagtt ccttgtatca aaagaaaaga gttgaactga aaatttcagt 960
atatacacia ttataatagc taggtgatta tttcatt 997

```

&lt;210&gt; 27

&lt;211&gt; 1481

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3255641CB1

&lt;400&gt; 27

```

ctgatccggt tcttgggtgcc cctggggcatc accaacatag ccatcgactt cggggagcag 60
gccttgaacc ggggcattgc tgctgtcaag gaggatgcag tcgagatget ggccagctac 120
gggctggcgt actccctcat gaagttcttc acgggtccca tgagtgaactt caaaaatgtg 180
ggcctgggtgt ttgtgaacag caagagagac aggaccaaag ccgtcctgtg tatggtggtg 240
gcagggggcca tcgctgccgt ctttcacaca ctgatagctt atagtgaattt aggatactac 300
attatcaata aactgcacca tgtggacgag tcggtgggga gcaagacgag aagggccttc 360
ctgtacctcg ccgcctttcc tttcatggac gcaatggcat ggacccatgc tggcattctc 420
ttaaaacaca aatacagttt cctgggtggga tgtgectcaa tctcagatgt catagctcag 480
gttgtttttg tagccatttt gcttcacagt cacctggaat gccgggagcc cctgctcatc 540
ccgacccctc ccttgtacat gggcgcaact gtgcgctgca ccaccctgtg cctgggctac 600
tacaagaaca ttacgacat catccctgac agaagtggcc cggagctggg gggagatgca 660
acaataagaa agatgctgag cttctgggtg cctttggctc taattctggc cacacagaga 720
atcagtcggc ctattgtcaa cctctttgtt tcccgggacc ttggtggcag ttctgcagcc 780
acagaggcag tggcgatttt gacagccaca tacctgtgtg gtcacatgcc atacggctgg 840
ttgacggaaa tccgtgctgt gtatcctgct ttogacaaga ataaccacag caacaaactg 900
gtgagcacga gcaacacagt cacggcagcc cacatcaaga agttcacctt cgtctgcatg 960
gctctgtcac tcacgctctg tttcgtgatg ttttggacac ccaacgtgtc tgagaaaatc 1020
ttgatagaca tcacggaggt ggactttgcc tttgcagaac tctgtgttgt tcctttgcgg 1080
atcttctcct tcttcccagt tccagtcaca gtgagggcgc atctcacccg gtggctgatg 1140
acactgaaga aaaccttcgt ccttgccccc agctctgtgc tgcggatcat cgtcctcatc 1200
gccagcctcg tggctctacc ctacctgggg gtgcacgggt cgaccctggg cgtgggctcc 1260
ctcctggcgg gctttgtggg agaatccacc atgggtcgcca tcgctgcgtg ctatgtctac 1320
cggaagcaga aaaagaagat ggagaatgag tcggccacgg agggggaaga ctctgccatg 1380
acagacatgc ctccgacaga ggaggtgaca gacatcgtgg aaatgagaga ggagaatgaa 1440
taaggcacgg gacgccatgg gcactgcagg gacagtcagt c 1481

```

&lt;210&gt; 28

&lt;211&gt; 303

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3620391CB1

<400> 28  
 tctagagccg ccgctgccc ggtctggagc gcgcgctccg ccgaggacaa gaccctggcc 60  
 tcacgccgga gcagcccat catgccgagg gagcgaggg agcgggatgc gaaggagcgg 120  
 gacaccatga aggaggacgg cggcgcgagg ttctcggtc gctccaggaa gaggaaggca 180  
 aacgtgacgg ttttttgcag gatccagatg aagaaatggc caaaatcgac aggacggcga 240  
 tggaccagtg tgggagccag acttgggaga atgatgcagt ctgtgcaggc ccctgtctcc 300  
 tga 303

<210> 29  
 <211> 1452  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 3969860CB1

<400> 29  
 gcgctctaat gttaccactc tgcctcctgt tctaggcttt tctggctttg tctgcccgga 60  
 tgcccagtgc cctgagaagg agctacatgc cacgggctgg aggctagagc ctgtggcacc 120  
 atgatttgcc tccgatggag gtggctgaat taggcttccc agagactgca gtgtcccaat 180  
 ccaggatctg tctatgtgct gtattgtgtg gccactggga ctttgcagac atgatggtga 240  
 taaggagcct gagtggacat ggctgcactc ttccaagaag caagcagctg tcccgtctgc 300  
 tcagactatc tggaaaaacc aatgtccctg gagtgtggat gcgcgctctg cctcaagtgc 360  
 attaatccac tgcagaagga gcccctatggg gaggatctac tttgctgttg ctcttccatg 420  
 gtctctcgga agaacaaaat caggcgcaat cggcagctag agaggtggc tcccacatc 480  
 aaggaaactgg agcccaagct gaagaagata ctgcagatga acccaaggat gcggaagtgc 540  
 caagtggata tgaccttggg tgccaacaca gccaacaact tcctcctcat ttctgacgac 600  
 ctcaaggagcg tccgaagtgg gcgcctcaga cagaatcggc aagaccttgc cgagagattt 660  
 gacgtgtccg tttgcatcct gggctccccct cgctttacct gtggccgcca ctgctgggag 720  
 gtggacgtgg gaacaagcac agaattgggac ctgggagtct gcagagaatc tgttcaccgc 780  
 aaaggaggga tccagctgac cacagagctt ggattctgga ctgtgagttt gagggatgga 840  
 ggccgcctct ctgccagcac ggtgccgctg actttcctct tcgtagaccg caagttacag 900  
 cgagtgggga tttttctgga tatgggcatg cagaacgttt ctttttttga tgctgaagt 960  
 ggttcccatg tctatacatt caggagcgtg tctgctgagg agccattgag cccatttttg 1020  
 gtccttcag ttcacctaata tggatgacaa ggtgtcttga gcctctgtcc tttgatgaac 1080  
 tcaggcacta ctgatgtccc agtcgctcct ggggaggcca aataagccct cactccaaaa 1140  
 aaacaaaaaa cagggttaaga aaattacttg ggtgggtaga cttagggaacg ctctacttcg 1200  
 taaaagcatt atacaaagtc acgggagaaa aatatgggac atttcttgat tgtacttaat 1260  
 ctaatttgat tagattatag agtcctaagt attaattatt gccaccatca aactcattga 1320  
 gtcctatggg tcacatcttg tttcctatag aaatgtcctg tattctggga tcaatttcca 1380  
 aatgctttac ttttttatct ctgcaagtgc aaattaatgt attatagaag ttatgagtta 1440  
 aatagaagag ta 1452

<210> 30  
 <211> 495  
 <212> DNA  
 <213> Homo sapiens



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4286006CB1

&lt;400&gt; 30

```

gtttgtctct caagttaaac caacaagccg atagaaaaag gtagttatca agagattttt 60
aaaacttcaa ccctttttct cttatagtta gtgaagagag tagaatatct ccagtttttg 120
ctgacatctc tacaacctga acaattggct taaacttcac ttgggattcc cggttgcttg 180
ttttagcatg gcgaaatttg gcgttcacag aatccttctt ctggctattt ctctgacaaa 240
gtgtctggag agtacaaaac tgctggcaga ccttaaaaaa tgtggtgact tggaaatgtga 300
agctttaata aacagagtct cagccatgag agattataga ggacctgact gccgatacct 360
gaacttcaact aaggggagaag agatatctgt ttatgttaaa cttgcaggag ataggggaaga 420
tttgtgggca ggaagtaaag gaaaggagtt tggatatttt cccagagatg cagtccagat 480
ttgagagggg gtcag                                     495

```

&lt;210&gt; 31

&lt;211&gt; 1993

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4325626CB1

&lt;400&gt; 31

```

gccgcgtacg gtgtgacctt tagacaattc tgtctcacag gatggacgtg gtagaggtcg 60
cgggcagttg gtgggcacaa gagcgagagg acatcattat gaaatacga aagggaacacc 120
gagctgggct gccagaggac aaggggccta agccttttcg aagctacaac aacaacgtcg 180
atcatttggg gattgtacat gagacggagc tgctcctct gactgcgcgg gaggcgaagc 240
aaattcggcg ggagatcagc cgaaagagca agtgggtgga tatgctggga gactgggaga 300
aatacaaaaag cagcagaaaag ctcatagatc gagcgtacaa gggaatgcc atgaacatcc 360
ggggcccgat gtggtcagtc ctcttgaaca ctgaggaaat gaagttgaaa aaccccgga 420
gataccagat catgaaggag aagggcaaga ggtcatctga gcacatccag cgcacgcacc 480
gggacgtaag cgggacatta aggaagcata tattcttcag ggatcgatac ggaaccaagc 540
agcgggaact actccacatc ctctggcat atgaggagta taacccggag gtgggctact 600
gcagggacct gagccacatc gccgccttgt tctctctta tcttctgag gaggatgcat 660
tctgggcact ggtgcagctg ctggccagtg agaggcact cctgcaggga ttccacagcc 720
caaattggcg gaccgtccag gggctccaag accaacagga gcatgtggtg gccacgtcac 780
aacccaagac catggggcat caggacaaga aagatctatg tgggcagtg tccccgttag 840
gctgcctcat ccgatattg attgacggga tctctctcgg gctcacccctg cgctgtggg 900
acgtgtatct ggtagaaggc gaacaggcgt tgatgccgat aacaagaatc gcctttaagg 960
ttcagcagaa ggcctcacg aagacgtcca ggtgtggccc gtgggcacgt ttttgcaacc 1020
ggttcgttga tacctgggccc agggatgagg acactgtgct caagcatctt agggcctcta 1080
tgaagaaaact aacaagaaag cagggggacc tgccaccccc agccaaacc gagcaagggt 1140
cgtcggcatc caggcctgtg ccggcttcac gtggcgga gacctctgc aagggggaca 1200
ggcaggcccc tcaggccca ccagcccggt tcccgcgcc catttggtca gcttccccgc 1260
cacgggcacc tcgttcttcc acacctgtc ctgggtgggc tgctcggga gacacctacc 1320
ctgtgggcac tcagggtgtg cccagcccg cctggctca gggaggacct cagggttct 1380
ggagattcct gcagtgaac tccatgcccc gcctcccaac ggacctggac gtagagggcc 1440
cttggttccg ccattatgat ttcagacaga gctgctgggt ccgtgccata tccaggagg 1500
accagctggc cccctgctgg caggctgaac accctgcgga gcgggtgaga tcggctttcg 1560
ctgcacccag cactgattcc gaccagggca ccccttcag agctagggac gaacagccgt 1620
gtgctccac ctcagggcct tgctctcgc gcctccactt ggaaagtct cagttccctc 1680
caggcttcta gaagcatctg ggccagggt catggctgga taatttcctt aggtttaaca 1740

```

```

acccaagcaa gttcgcatcc tcgttttatt tttgggttaa cttatgaaaa tgtattaaga 1800
aagagtgcag ctcgagagag attcagagat ggaacacacc agaccccaga tcacaaagcc 1860
aaccatgccc agccctcccc agcaccacca gcccacgac catcgttctg aattctgacg 1920
acaccgtgag cctgcctttg tacttcaaac tcatggaagg ataaccacct tcatgttttg 1980
aaataaatgg gtc                                     1993

```

&lt;210&gt; 32

&lt;211&gt; 728

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1438978CB1

&lt;400&gt; 32

```

aagctcttcc attgggctgt tgagtgggtgg ccgaccgacg gccaggagtt tcttttctgc 60
gcttggtcgt tttctgttcg gtttccttcc cgctagcggg gccacgaggg ttgctaggca 120
acagccctcg ggtgacttgg tcttaggggc ctgtccggct tggggctgat gaaaggagct 180
gtccgcgccc gggctcttcc gagaagtggg tgctgacagc cacaaagtga aagggagtga 240
ggcggcgtgg acgagtaagg agtgacagtg aggattcaca tttgggttat ttcaagatga 300
gcttctact gcccaagctg actagcaaaa aggaagtaga ccaggcgata aaaagtactg 360
ctgagaaggt gttggttctc aggtttggga gagatgaaga tctgtctgt ctgcagctag 420
atgatattct ttctaagacc tcttctgact taagtaaaat ggctgctata tacctggtag 480
atgtggacca aactgcagtt tatacacagt attttgacat cagttatatt ccactactg 540
tcttttctt caatgggcag catatgaaag tggattatgg gtaagtgcag ttgatctgaa 600
gttaattgca acctgtgaag ttctcttggg aagcattttc agtagcttgc ctatttccat 660
gtgatgttgg ctctgtgagt cttatatcag tactgtttcc tcaattgacg cactctctaa 720
ttttttat                                     728

```

&lt;210&gt; 33

&lt;211&gt; 1452

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2024773CB1

&lt;400&gt; 33

```

gtcaggagcg tagaggcggc ggcaaaatgg cggcgccctga ggagcgggat ctaaccacagg 60
agcagacaga gaagctgctg cagtttcagg atctcactgg catcgaatct atggatcagt 120
gtcgccatac cttggaacag cataactgga acatagaggc tgctgtacag gacagattga 180
atgagcaaga gggcgctacct agtgttttca acccacctcc atcacgaccc ctgcaggtta 240
atacagctga ccacaggatc tacagctatg ttgtctcaag acctcaacca agggggctgc 300
ttggatgggg ttattacttg ataagtctc cattccggtt tacctattac acgatacttg 360
atatatttag gtttgcctct cgttttatac ggctgaccc tcgcagccgg gtcactgacc 420
ccgttgggga cattgtttca tttatgcact cttttgaaga gaaatatggg agggcacacc 480
ctgtcttcta ccagggaacg tacagccagg cacttaacga tgccaaaagg gagcttcgct 540
ttcttttggg ttatcttcat ggagatgatc accaggactc tgatgagttt tgtcgcaaca 600
cactctgtgc acctgaagtt atttcactaa taaacactag gatgctcttc tgggcatgct 660
ctacaaacaa acctgaggga tacagggtct cacaggcttt acgagagaac acctatccat 720
tcctggccat gattatgctg aaggatcgaa ggatgactgt ggtgggacgg ctagaaggcc 780

```

```

tcattcaacc tgatgacctc attaaccaac tgacatttat catggatgct aaccagactt 840
acctgggtgtc agaacgccta gaaaggggaag aaagaaacca gaccaagtgt ctgagacaac 900
agcaggatga ggcctacctg gectctctca gagctgacca ggagaaagaa agaaagaaac 960
gggaggagcg ggagcgtaag cggcggaagg aggaggagggt gcaacagcaa aagttggcag 1020
aggagagacg gcggcagaat ttacaggagg aaaaggaaaag gaagttggaa tgcctgcccc 1080
ctgaaccttc ccctgatgac cctgaaagtgt tcaagatcat cttcaaatta cctaattgatt 1140
ctcgagtaga gagacgattc caotttttcac agtctctaac agtaatccac gacttcttat 1200
tctccttgaa ggaaagccca gaaaagtttc agattgaagc caattttccc aggcgagtgc 1260
tgccctgcat cccttcagag gagtggccca atccccctac gctacaggag gccggactca 1320
gccacacaga agttcttttt gttcaggacc taactgacga atgacatttt tttcttctgt 1380
tccccctcta cccagtcctc taaaagaaat ggggaaaaaa gaaaacaaca gcaagtcaaa 1440
aaaaaaaaaa aa 1452

```

&lt;210&gt; 34

&lt;211&gt; 1229

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3869790CB1

&lt;400&gt; 34

```

ccaggcttga ctcatccca ccttgtcctg ggctgagatc ccaggtttgt aacagaaaac 60
accactaaag cccagcaca ggagagaacc acccagccca gaagttccag ggaagggaact 120
ctccgggtcca ccatggagta cctctcagct ctgaacccca gtgacttact cagggtcagta 180
tctaataataa gctcggagtt tggacggagg gtctggacct cagctccacc accccagcga 240
cctttccgtg tctgtgatca caagcggacc atccggaaaag gctgacagc tgccaccgcg 300
caggagctgc tagccaaagc attggagacc ctactgctga atggagtgt aaccttggtg 360
ctagaggagg atggaactgc agtggacagt gaggacttct tccagctgt ggaggatgac 420
acgtgcctga tgggtgttgc gtctgggtcag agctggagcc ctacaaggag tggagtgtgt 480
tcatatggcc tgggacggga gaggcccaag cacagcaagg acatcgcccg attcaccttt 540
gacgtgtaca agcaaaaccc tcgagacctc tttggcagcc tgaatgtcaa agccacattc 600
tacgggtctt actctatgag ttgtgacttt caaggacttg gcccaaagaa agtactcagg 660
gagctccttc gttggacctc cacactgctg caaggcctgg gccatatgtt gctgggaatt 720
tcctccaccc ttcgtcatgc agtggagggg gctgagcagt ggcagcagaa ggccgcctc 780
cattcctact aaggggtctt gagcttctgc cccagaatc attccaaccg acccactgca 840
aagactatga cagcatcaaa tttcaggacc tgcagacagt acaggctaga taaccacccc 900
aatttcccca ctgtcctctg atccccctgt gacagaacct ttcagcataa cgcctcacat 960
cccaagtcta tacccttacc tgaagaatgc tgttctttcc tagccacctt tctagcctcc 1020
cacttgccct gaaaggccaa gatcaagatg tccccaggc atcttgatcc cagectgact 1080
gctgctacat ctaatccctt accaatgcct cctgtcccta aactccccag catactgatg 1140
acagccctct ctgactttac cttgagatct gtcttcatac ctttccctc aaactaacia 1200
aaacatttcc aataaaaata tcagaatac 1229

```

&lt;210&gt; 35

&lt;211&gt; 1455

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 001273CB1

&lt;400&gt; 35

```

tggaaatcgcg ggcaaagatg ggggcccga ggtgttgag gcctttgcta cgcggtccga 60
ggctttcatt gcacaccgcg gctaataccg ccgccacggc tacagaaacg acctgccaaag 120
acgtcgcggc gacccccgtc ggcgggtacc cgcgattgt ggctccatg acagccgaca 180
gcaaagctgc acggctgcgg cggatcgagc gctggcaggc gacggtgcac gctgcgagtg 240
cggtagacga gaagctgcga atcctcacca agatgcagtt tatgaagtac atggtttacc 300
cgcagacctt cgcgctgaat gccgaccgct ggtaccagta cttaccaag accgtgttcc 360
tgtcgggtct gccgcccgcg ccagcgagc ccgagcccga gccgaaacc gaacctgaac 420
ctgcgctgga cctcgcggcg ctgcgtgcgg tcgcctgcga ctgcctgctg caggagact 480
tctacctgcg gcgcaggcg cgctgcacc gttacgagga gagcgaggc atatctttgc 540
ccttctgga tcagctggtg tcaacctcg tgggcctcct cagcccacac aaccggccc 600
tggcgcgtgc cgcctcgat tatagatgcc cagttcattt ttactgggtg cgtggtgaag 660
aaattattcc tcgtggtcat cgaagaggc gaattgatga cttgcgatac cagatagatg 720
ataaaccaaa caaccagatt cgaatatcca agcaactcgc agagtttgtg ccattggatt 780
attctgttcc tatagaaatc cccactataa aatgtaaacc agacaaactt ccattattca 840
aacggcagta tgaaaaccac atatttgttg gtcaaaaaac tgcagatcct tgctgttacg 900
gtcacaccca gtttcatctg ttacctgaca aattaagaag ggaaaggctt ttgagacaaa 960
actgtgctga tcagatagaa gttgttttta gagctaatgc tattgcaagc ctttttgctt 1020
ggactggagc acaagctatg tatcaaggat tctggagtga agcagatgtt actcgacctt 1080
ttgtctccca ggtgtgata caagatggaa aatacttttc ctttttctgc taccagtaa 1140
tacttttggc actgactaca caagctgac aaaataacc tcgtaaaaat atatgttggg 1200
gtacacaaag taagcctctt tatgaaacaa ttgaggataa tgatgtgaaa ggttttaagt 1260
atgatgttct acttcagata gttcactttc tactgaatag accaaaagaa gaaaaatcac 1320
agctgttggg aaactgaaaa agcatatttg attgagaact gtgggaatat taaattttta 1380
ctgaagggaac aataatgatg agatttgtta ctgtcaacta ttaaatacat tgatttttga 1440
gacaaaaaaa aaaaa 1455

```

&lt;210&gt; 36

&lt;211&gt; 2099

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 411831CB1

&lt;400&gt; 36

```

gaacgctggt tagtcttgtt tcgggttccg gctgcgttg gcttgcgtgc ggctcgctaa 60
gactatggcg tccgggcctc attcgacagc tactgctgcc gcagccgcct catcgccgc 120
cccaagcgcg ggcggctcca gctccgggac gacgaccacg acgacgacca cgacgggagg 180
gatectgatc ggcgatcgcc tgtactcgga agtttcaactt accatcgacc actctctgat 240
tccggaggag aggtctctgc ccaccccatc catgcaggat gggctcgacc tgcccagtga 300
gacggactta cgcacctggt gctgcgagct catccaggcc gccggcatte tcctccggct 360
gccgcaggtg gcgatggcaa cggggcagggt gttgtttcat cgttttttct actccaaatc 420
tttcgtcaaa cacagtttcg agattgttgc tatggcttgt attaatcttg catcaaaaat 480
cgaagaagca cctagaagaa taagagatgt gattaatgta ttccaccacc tccgccagtt 540
aagaggaaaa aggactccaa gcccctgat ccttgatcag aactacatta acacaaaaaa 600
tcaagttatc aaagcagaga ggagggtgct aaaggagtgt ggattttgtg ttcatgtcaa 660
gcacctcatc aagatcattg ttatgtattt acaagtctta gaatgtgaac gtaatcaaac 720
cctggttcaa actgcctgga attacatgaa tgacagtctt cgaaccaatg tgtttgttcg 780
atttcaacca gagactatag catgtgcttg catctacctt gcagctagag cacttcagat 840
tccgttgcca actcgctccc attgggttct tcttttttgt actacagaag aggaaatcca 900
ggaaatctgc atagaaacac ttaggcttta taccagaaaa aagccaaact atgaattact 960
ggaaaaagaa gtagaaaaaa gaaaagtagc cttacaagaa gccaaattaa aagcaaaggg 1020

```

```

attgaatccg gatggaactc cagccctttc aaccctgggt ggattttctc cagcctccaa 1080
gccatcatca ccaagagaag taaaagctga agagaaatca ccaatctcca ttaatgtgaa 1140
gacagtcaaa aaagaacctg aggatagaca acaggcttcc aaaagccctt acaatggtgt 1200
aagaaaagac agcaagagaa gtagaaatag cagaagtgc aatcgatcga ggtcaagaac 1260
acgatcacgt tctagatcac atactccaag aagacactat aataataggc ggagtcgatc 1320
tggaacatac agctcgagat caagaagcag gtcccgagc cacagtgaag gccctogaag 1380
acatcataat catggtttct ctcaccttaa ggccaagcat accagagatg atttaaaaaag 1440
ttcaaacaga catggtcata aaaggaaaaa atctcgttct cgatctcaga gcaagtctcg 1500
ggatcactca gatgcagcca agaaacacag gcatgaaagg ggacatcata gggacaggcg 1560
tgaacgatct cgctcctttg agagggtccc taaaagcaag caccatggtg gcagtcgctc 1620
aggacatggc aggcacaggc gctgactttc tcttcctttg agcctgcac agttcttggg 1680
tttgccatc tacagtgtga tgtatggact caatcaaaaa cattaaacgc aaactgatta 1740
ggatttgatt tcttgaaacc ctctaggtct ctagaacact gaggacagtt tcttttgaaa 1800
agaactatgt taattttttt gcacattaaa atgccctagc agtatctaataaaaaacat 1860
ggtcagggtc aattgtactt tattatagtt gtgtattgtt tattgctata agaactggag 1920
cgtgaattct gtaaaaatgt atcttatttt tatacagata aaattgcaga cactgttcta 1980
tttaagtggg tatttgttta aatgatgggt aatactttct taacactggg ttgtctgcat 2040
gtgtaaagat ttttacaagg aaataaaata caaatcttgt tttttctaaa aaaaaaaaaa 2099

```

<210> 37

<211> 1363

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1520835CB1

<400> 37

```

gacccagag gccaccctgg ccacttccag aaagctgtgg gccctgggat actcccctcc 60
caggggtgtc ggtggcaggc ctgtgectat ccctgctgtc cccaggggtg gcccggggg 120
tcaggagctc cagaagggcc agctgggcat attctgagat tggccatcag ccccatttc 180
tgctgcaaac ctggtcagag ccagtgttcc cctcatgggg cctaaagaca gtgccaagt 240
cctgcaccgt ggaccacagc cgagccactg gccagccggg gatggtccca cgcaggagcg 300
ctgtggacce cgctctctgg gcagccctgt cctaggcctg gacacctgca gagcctggga 360
ccacgtggat gggcagatcc tgggccagct gcggccctg acagaggagg aagaggagga 420
gggcgcggg gccacctgt ccagggggcc tgccttcccc ggcatgggct ctgaggagtt 480
gcgtctggcc tccttctatg actggccgct gactgctgag gtgccaccg agctgctggc 540
tgctgccggc ttcttccaca caggccatca ggacaagggt aggtgcttct tctgctatgg 600
gggcctgcag agctggaagc gcggggacga cccctggacg gagcatgcca agtggttccc 660
cagctgtcag ttctgtctcc ggtcaaaagg aagagacttt gtccacagtg tgcaggagac 720
tcactcccag ctgctgggct cctgggaccc gtgggaagaa ccggaagacg cagccctgt 780
ggcccccctc gtccttgcct ctgggtaccc tgagctgccc acaccagga gagaggcca 840
gtctgaaagt gccaggagc caggagggt cagtccagcc gagggccaga gggcgtgggt 900
gaggaactgc aagggtgtcc tggaccgcgc cgtgtccatc gtctttgtgc cgtgcggcca 1020
cctggtctgt gctgagtgtg ccccgccct gcagctgtgc cccatctgca gagccccct 1080
ccgcagccgc gtgcgcacct tcctgtccta ggccagggtc catggccggc cagggtgggt 1140
gcagagtggg ctccctgccc ctctctgect gttctggact gtgttctggg cctgctgagg 1200
atggcagagc tgggtgtccat ccagcactga ccagccctga tccccgacc accgcccagg 1260
gtggagaagg aggccttgc ttggcgtggg ggatggctta actgtacctg tttggatgct 1320
tctgaataga aataaagtgg gttttccctg gaaaaaaaaa aaa 1363

```

<210> 38  
 <211> 1465  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1902803CB1

<400> 38  
 ggggttaaata caaagcaagg agaatgaaaa ggaccctccc tctgaaatac agcaccggcc 60  
 tcagcatgga ggcaaaagga cagtgggcaa gaccacgtcc ctccgaaggg agatggctgc 120  
 ggggatgtat ttggaacatt atctggacag tattgaaaac cttccctttg aattacagag 180  
 aaactttcag ctcagtaggg acctagacca aagaacagag gacctgaagg ctgaaattga 240  
 caagttggcc actgagtata tgagtagtgc ccgcagcctg agctccgagg aaaaattggc 300  
 ccttctcaaa cagatccagg aagcctatgg caagtgcagg gaatttggtg acgacaaggt 360  
 gcagcttgcc atgcagacct atgagatggt ggacaaacac attcggcgcc tggacacaga 420  
 cctggcccgt tttgaggctg atctcaagga gaaacagatt gagtcaagt actatgacag 480  
 ctcttccagc aaaggcaaaa agaaaggccg gactcaaaag gagaagaaag ctgctcgtgc 540  
 tcgttccaaa gggaaaaaact cggatgaaga agcccccaag actgcccaga agaagttaaa 600  
 gctcgtgcgc acaagtcctg agtatgggat gccctcagt acctttggca gtgtccaccc 660  
 ctctgatgtg ttggatatgc ctgtggatcc caacgaaccc acctattgcc tttgtcacca 720  
 ggtctcctat ggagagatga ttggctgtga caaccctgat tgttccattg agtggttcca 780  
 ttttgccctg ttggggctga caaccaagcc tcgggggaaa tggttttgcc cacgctgctc 840  
 ccaagaacgg aagaagaaat agataagggc cttggattcc aacacagttt cttccacatc 900  
 ccctgacttg ggctagtggg cagaggaatg cctgtgctgg ggccaggggt tcagggagga 960  
 gtggatggca cagtgtctgc atcccttctc ctccctctc cccactccc gtgctgaggc 1020  
 tgcacagac cctggtaggg aggggtgccc cagccactaa cggatgtgc tctccttcag 1080  
 ccctctccct tcggagggac gtgtcttgc ccactgtcct tttgctcca tgcagaggc 1140  
 ggtgctgtat ttcagaggga gggctctttt cattctcctt gctttgtatt taaggactgg 1200  
 ggcatagcat gggggcagtc ccccagacct cttcattccc cctcctgtgg tgagggctag 1260  
 gtgtgatcaa cacttttctt ctccattccc ttctgtctt tttcatgggt ggggatccac 1320  
 caggatcatc aggcctctggc cctagttgaa ggggcacccc ttctctgtg ccaagaggat 1380  
 tcactctggg agagggggca aggtggaatg cagataactc acatgtaaaa ggaacttggg 1440  
 taggtaaata aaagctatac atggt 1465

<210> 39  
 <211> 332  
 <212> PRT  
 <213> Mus musculus

<300>  
 <308> GenBank ID No: g452276

<400> 39  
 Met Ala Thr Pro Val Pro Pro Pro Ser Pro Arg His Leu Arg Leu  
 1 5 10 15  
 Leu Arg Leu Leu Leu Ser Gly Leu Ile Leu Gly Ala Ala Leu Asn  
 20 25 30  
 Gly Ala Thr Ala Arg Arg Pro Asp Ala Thr Thr Cys Pro Gly Ser

	35		40		45
Leu Asp Cys Ala	Leu Lys Arg Arg Ala Lys Cys Pro Pro Gly Ala				
	50		55		60
His Ala Cys Gly	Pro Cys Leu Gln Ser Phe Gln Glu Asp Gln Arg				
	65		70		75
Gly Phe Cys Val	Pro Arg Lys His Leu Ser Ser Gly Glu Gly Leu				
	80		85		90
Pro Gln Pro Arg	Leu Glu Glu Glu Ile Asp Ser Leu Ala Gln Glu				
	95		100		105
Leu Ala Leu Lys	Glu Lys Glu Ala Gly His Ser Arg Leu Thr Ala				
	110		115		120
Gln Pro Leu Leu	Glu Arg Ala Gln Lys Leu Leu Glu Pro Ala Ala				
	125		130		135
Thr Leu Gly Phe	Ser Gln Trp Gly Gln Arg Leu Glu Pro Gly Leu				
	140		145		150
Pro Ser Thr His	Gly Thr Ser Ser Pro Ile Pro His Thr Ser Leu				
	155		160		165
Ser Ser Arg Ala	Ser Ser Gly Pro Val Gln Met Ser Pro Leu Glu				
	170		175		180
Pro Gln Gly Arg	His Gly Asn Gly Leu Thr Leu Val Leu Ile Leu				
	185		190		195
Ala Phe Cys Leu	Ala Ser Ser Ala Ala Leu Ala Val Ala Ala Leu				
	200		205		210
Cys Trp Cys Arg	Leu Gln Arg Glu Ile Arg Leu Thr Gln Lys Ala				
	215		220		225
Asp Tyr Ala Ala	Thr Ala Lys Gly Pro Thr Ser Pro Ser Thr Pro				
	230		235		240
Arg Ile Ser Pro	Gly Asp Gln Arg Leu Ala His Ser Ala Glu Met				
	245		250		255
Tyr His Tyr Gln	His Gln Arg Gln Gln Met Leu Cys Leu Glu Arg				
	260		265		270
His Lys Glu Pro	Pro Lys Glu Leu Glu Ser Ala Ser Ser Asp Glu				
	275		280		285
Glu Asn Glu Asp	Gly Asp Phe Thr Val Tyr Glu Cys Pro Gly Leu				
	290		295		300
Ala Pro Thr Gly	Glu Met Glu Val Arg Asn Pro Leu Phe Asp His				
	305		310		315
Ser Thr Leu Ser	Ala Pro Val Pro Gly Pro His Ser Leu Pro Pro				
	320		325		330
Leu Gln					

&lt;210&gt; 40

&lt;211&gt; 268

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;300&gt;

&lt;308&gt; GenBank ID No: g998357

&lt;400&gt; 40

Met Ala Val Asn Val Tyr Ser Thr Ser Val Thr Ser Asp Asn Leu
1 5 10 15

```

Ser Arg His Asp Met Leu Ala Trp Ile Asn Glu Ser Leu Gln Leu
      20                      25                      30
Asn Leu Thr Lys Ile Glu Gln Leu Cys Ser Gly Ala Ala Tyr Cys
      35                      40                      45
Gln Phe Met Asp Met Leu Phe Pro Gly Ser Ile Ala Leu Lys Lys
      50                      55                      60
Val Lys Phe Gln Ala Lys Leu Glu His Glu Tyr Ile Gln Asn Phe
      65                      70                      75
Lys Ile Leu Gln Ala Gly Phe Lys Arg Met Gly Val Asp Lys Ile
      80                      85                      90
Ile Pro Val Asp Lys Leu Val Lys Gly Lys Phe Gln Asp Asn Phe
      95                      100                     105
Glu Phe Val Gln Trp Phe Lys Lys Phe Phe Asp Ala Asn Tyr Asp
     110                      115                     120
Gly Lys Asp Tyr Asp Pro Val Ala Ala Arg Gln Gly Gln Glu Thr
     125                      130                     135
Ala Val Ala Pro Ser Leu Val Ala Pro Ala Leu Asn Lys Pro Lys
     140                      145                     150
Lys Pro Leu Thr Ser Ser Ser Ala Ala Pro Gln Arg Pro Ile Ser
     155                      160                     165
Thr Gln Arg Thr Ala Ala Ala Pro Lys Ala Gly Pro Gly Val Val
     170                      175                     180
Arg Lys Asn Pro Gly Val Gly Asn Gly Asp Asp Glu Ala Ala Glu
     185                      190                     195
Leu Met Gln Gln Val Asn Val Leu Lys Leu Thr Val Glu Asp Leu
     200                      205                     210
Glu Lys Glu Arg Asp Phe Tyr Phe Gly Lys Leu Arg Asn Ile Glu
     215                      220                     225
Leu Ile Cys Gln Glu Asn Glu Gly Glu Asn Asp Pro Val Leu Gln
     230                      235                     240
Arg Ile Val Asp Ile Leu Tyr Ala Thr Asp Glu Gly Phe Val Ile
     245                      250                     255
Pro Asp Glu Gly Gly Pro Gln Glu Glu Gln Glu Glu Tyr
     260                      265

```

&lt;210&gt; 41

&lt;211&gt; 418

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;300&gt;

&lt;308&gt; GenBank ID No: g455719

&lt;400&gt; 41

```

Met Gly Glu Asp Ala Ala Gln Ala Glu Lys Phe Gln His Pro Asn
  1                      5                      10                      15
Thr Asp Met Leu Gln Glu Lys Pro Ser Ser Pro Ser Pro Met Pro
      20                      25                      30
Ser Ser Thr Pro Ser Pro Ser Leu Asn Leu Gly Ser Thr Glu Glu
      35                      40                      45
Ala Ile Arg Asp Asn Ser Gln Val Asn Ala Val Thr Val His Thr
      50                      55                      60

```



Leu	Leu	Asp	Lys	Leu	Val	Asn	Met	Leu	Asp	Ala	Val	Arg	Glu	Asn	
				65					70					75	
Gln	His	Asn	Met	Glu	Gln	Arg	Gln	Ile	Asn	Leu	Glu	Gly	Ser	Val	
				80					85					90	
Lys	Gly	Ile	Gln	Asn	Asp	Leu	Thr	Lys	Leu	Ser	Lys	Tyr	Gln	Ala	
				95					100					105	
Ser	Thr	Ser	Asn	Thr	Val	Ser	Lys	Leu	Leu	Glu	Lys	Ser	Arg	Lys	
				110					115					120	
Val	Ser	Ala	His	Thr	Arg	Ala	Val	Arg	Glu	Arg	Leu	Glu	Arg	Gln	
				125					130					135	
Cys	Val	Gln	Val	Lys	Arg	Leu	Glu	Asn	Asn	His	Ala	Gln	Leu	Leu	
				140					145					150	
Arg	Arg	Asn	His	Phe	Lys	Val	Leu	Ile	Phe	Gln	Glu	Glu	Ser	Glu	
				155					160					165	
Ile	Pro	Ala	Ser	Val	Phe	Val	Lys	Glu	Pro	Val	Pro	Ser	Ala	Ala	
				170					175					180	
Glu	Gly	Lys	Glu	Glu	Leu	Ala	Asp	Glu	Asn	Lys	Ser	Leu	Glu	Glu	
				185					190					195	
Thr	Leu	His	Asn	Val	Asp	Leu	Ser	Ser	Asp	Asp	Glu	Leu	Pro	Arg	
				200					205					210	
Asp	Glu	Glu	Ala	Leu	Glu	Asp	Ser	Ala	Glu	Glu	Lys	Met	Glu	Glu	
				215					220					225	
Ser	Arg	Ala	Glu	Lys	Ile	Lys	Arg	Ser	Ser	Leu	Lys	Lys	Val	Asp	
				230					235					240	
Ser	Leu	Lys	Lys	Ala	Phe	Ser	Arg	Gln	Asn	Ile	Glu	Lys	Lys	Met	
				245					250					255	
Asn	Lys	Leu	Gly	Thr	Lys	Ile	Val	Ser	Val	Glu	Arg	Arg	Glu	Lys	
				260					265					270	
Ile	Lys	Lys	Ser	Leu	Thr	Pro	Asn	His	Gln	Lys	Ala	Ser	Ser	Gly	
				275					280					285	
Lys	Ser	Ser	Pro	Phe	Lys	Val	Ser	Pro	Leu	Ser	Phe	Gly	Arg	Lys	
				290					295					300	
Lys	Val	Arg	Glu	Gly	Glu	Ser	Ser	Val	Glu	Asn	Glu	Thr	Lys	Leu	
				305					310					315	
Glu	Asp	Gln	Met	Gln	Glu	Asp	Arg	Glu	Glu	Gly	Ser	Phe	Thr	Glu	
				320					325					330	
Gly	Leu	Ser	Glu	Ala	Ser	Leu	Pro	Ser	Gly	Leu	Met	Glu	Gly	Ser	
				335					340					345	
Ala	Glu	Asp	Ala	Glu	Lys	Ser	Ala	Arg	Arg	Gly	Asn	Asn	Ser	Ala	
				350					355					360	
Val	Gly	Ser	Asn	Ala	Asp	Leu	Thr	Ile	Glu	Glu	Asp	Glu	Glu	Glu	
				365					370					375	
Glu	Pro	Val	Ala	Leu	Gln	Gln	Ala	Gln	Gln	Val	Arg	Tyr	Glu	Ser	
				380					385					390	
Gly	Tyr	Met	Leu	Asn	Ser	Glu	Glu	Met	Glu	Glu	Pro	Ser	Glu	Lys	
				395					400					405	
Gln	Val	Gln	Pro	Ala	Val	Leu	His	Val	Asp	Gln	Thr	Ala			
				410					415						

&lt;210&gt; 42

&lt;211&gt; 142

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;300&gt;

&lt;308&gt; GenBank ID No: g2565275

&lt;400&gt; 42

```

Met Ser Tyr Met Leu Pro His Leu His Asn Gly Trp Gln Val Asp
 1              5              10              15
Gln Ala Ile Leu Ser Glu Glu Asp Arg Val Val Val Ile Arg Phe
              20              25              30
Gly His Asp Trp Asp Pro Thr Cys Met Lys Met Asp Glu Val Leu
              35              40              45
Tyr Ser Ile Ala Glu Lys Val Lys Asn Phe Ala Val Ile Tyr Leu
              50              55              60
Val Asp Ile Thr Glu Val Pro Asp Phe Asn Lys Met Tyr Glu Leu
              65              70              75
Tyr Asp Pro Cys Thr Val Met Phe Phe Phe Arg Asn Lys His Ile
              80              85              90
Met Ile Asp Leu Gly Thr Gly Asn Asn Asn Lys Ile Asn Trp Ala
              95              100             105
Met Glu Asp Lys Gln Glu Met Val Asp Ile Ile Glu Thr Val Tyr
              110             115             120
Arg Gly Ala Arg Lys Gly Arg Gly Leu Val Val Ser Pro Lys Asp
              125             130             135
Tyr Ser Thr Lys Tyr Arg Tyr
              140

```

&lt;210&gt; 43

&lt;211&gt; 464

&lt;212&gt; PRT

<213> *Drosophila melanogaster*

&lt;300&gt;

&lt;308&gt; GenBank ID NO: g3688609

&lt;400&gt; 43

```

Met Glu Ala Asp Gly Leu Thr Asn Glu Gln Thr Glu Lys Val Leu
 1              5              10              15
Gln Phe Gln Asp Leu Thr Gly Ile Glu Asp Met Asn Val Cys Arg
              20              25              30
Asp Val Leu Ile Arg His Gln Trp Asp Leu Glu Val Ala Phe Gln
              35              40              45
Glu Gln Leu Asn Ile Arg Glu Gly Arg Pro Thr Met Phe Ala Ala
              50              55              60
Ser Thr Asp Val Arg Ala Pro Ala Val Leu Asn Asp Arg Phe Leu
              65              70              75
Gln Gln Val Phe Ser Ala Asn Met Pro Gly Gly Arg Thr Val Ser
              80              85              90
Arg Val Pro Ser Gly Pro Val Pro Arg Ser Phe Thr Gly Ile Ile
              95             100             105
Gly Tyr Val Ile Asn Phe Val Phe Gln Tyr Phe Tyr Ser Thr Leu
              110             115             120
Thr Ser Ile Val Ser Ala Phe Val Asn Leu Gly Gly Gly Asn Glu
              125             130             135

```

Ala Arg Leu Val Thr Asp Pro Leu Gly Asp Val Met Lys Phe Ile	140	145	150
Arg Glu Tyr Tyr Glu Arg Tyr Pro Glu His Pro Val Phe Tyr Gln	155	160	165
Gly Thr Tyr Ala Gln Ala Leu Asn Asp Ala Lys Gln Glu Leu Arg	170	175	180
Phe Leu Ile Val Tyr Leu His Lys Asp Pro Ala Lys Asn Pro Asp	185	190	195
Val Glu Ser Phe Cys Arg Asn Thr Leu Ser Ala Arg Ser Val Ile	200	205	210
Asp Tyr Ile Asn Thr His Thr Leu Leu Trp Gly Cys Asp Val Ala	215	220	225
Thr Pro Glu Gly Tyr Arg Val Met Gln Ser Ile Thr Val Arg Ser	230	235	240
Tyr Pro Thr Met Val Met Ile Ser Leu Arg Ala Asn Arg Met Met	245	250	255
Ile Val Gly Arg Phe Glu Gly Asp Cys Thr Pro Glu Glu Leu Leu	260	265	270
Arg Arg Leu Gln Ser Val Thr Asn Ala Asn Glu Val Trp Leu Ser	275	280	285
Gln Ala Arg Ala Asp Arg Leu Glu Arg Asn Phe Thr Gln Thr Leu	290	295	300
Arg Arg Gln Gln Asp Glu Ala Tyr Glu Gln Ser Leu Leu Ala Asp	305	310	315
Glu Glu Lys Glu Arg Gln Arg Gln Arg Glu Arg Asp Ala Val Arg	320	325	330
Gln Ala Glu Glu Ala Val Glu Gln Ala Arg Arg Asp Val Glu Leu	335	340	345
Arg Lys Glu Glu Ile Ala Arg Gln Lys Ile Glu Leu Ala Thr Leu	350	355	360
Val Pro Ser Glu Pro Ala Ala Asp Ala Val Gly Ala Ile Ala Val	365	370	375
Val Phe Lys Leu Pro Ser Gly Thr Arg Leu Glu Arg Arg Phe Asn	380	385	390
Gln Thr Asp Ser Val Leu Asp Val Tyr His Tyr Leu Phe Cys His	395	400	405
Pro Asp Ser Pro Asp Glu Phe Glu Ile Thr Thr Asn Phe Pro Lys	410	415	420
Arg Val Leu Phe Ser Lys Ala Asn Leu Asp Ala Ala Gly Glu Thr	425	430	435
Gly Thr Ala Lys Glu Thr Leu Thr Lys Thr Leu Gln Ala Val Gly	440	445	450
Leu Lys Asn Arg Glu Leu Leu Phe Val Asn Asp Leu Glu Ala	455	460	

&lt;210&gt; 44

&lt;211&gt; 219

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;300&gt;

&lt;308&gt; GenBank ID No: g3114594

&lt;400&gt; 44

Met	Glu	Tyr	Leu	Ser	Ala	Phe	Asn	Pro	Asn	Gly	Leu	Leu	Arg	Ser	1	5	10	15
Val	Ser	Thr	Val	Ser	Ser	Glu	Leu	Ser	Arg	Arg	Val	Trp	Asn	Ser	20	25	30	
Ala	Pro	Pro	Pro	Gln	Arg	Pro	Phe	Arg	Val	Cys	Asp	His	Lys	Arg	35	40	45	
Thr	Val	Arg	Lys	Gly	Leu	Thr	Ala	Ala	Ser	Leu	Gln	Glu	Leu	Leu	50	55	60	
Asp	Lys	Val	Leu	Glu	Thr	Leu	Leu	Leu	Arg	Gly	Val	Leu	Thr	Leu	65	70	75	
Val	Leu	Glu	Glu	Asp	Gly	Thr	Ala	Val	Asp	Ser	Glu	Asp	Phe	Phe	80	85	90	
Gln	Leu	Leu	Glu	Asp	Asp	Thr	Cys	Leu	Met	Val	Leu	Glu	Gln	Gly	95	100	105	
Gln	Ser	Trp	Ser	Pro	Lys	Ser	Gly	Met	Leu	Ser	Tyr	Gly	Leu	Gly	110	115	120	
Arg	Glu	Lys	Pro	Lys	His	Ser	Lys	Asp	Ile	Ala	Arg	Ile	Thr	Phe	125	130	135	
Asp	Val	Tyr	Lys	Gln	Asn	Pro	Arg	Asp	Leu	Phe	Gly	Ser	Leu	Asn	140	145	150	
Val	Lys	Ala	Thr	Phe	Tyr	Gly	Leu	Tyr	Ser	Met	Ser	Cys	Asp	Phe	155	160	165	
Gln	Gly	Val	Gly	Pro	Lys	Arg	Val	Leu	Arg	Glu	Leu	Leu	Arg	Gly	170	175	180	
Thr	Ser	Ser	Gln	Leu	Gln	Gly	Leu	Gly	His	Met	Leu	Leu	Gly	Ile	185	190	195	
Ser	Ser	Thr	Leu	Arg	His	Val	Val	Glu	Gly	Ala	Asp	Arg	Trp	Gln	200	205	210	
Trp	His	Gly	Gln	Arg	His	Leu	His	Ser							215			



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, C07K 14/47, C12Q 1/68, A61K 38/17, C07K 16/18		A3	(11) International Publication Number: <b>WO 00/23589</b>
			(43) International Publication Date: 27 April 2000 (27.04.00)
(21) International Application Number: PCT/US99/24511		(72) Inventors; and	
(22) International Filing Date: 19 October 1999 (19.10.99)		(75) Inventors/Applicants (for US only): TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #12, Mountain View, CA 94040 (US). GUEGLER, Karl, J. [CH/US]; 1048 Oakland Avenue, Menlo Park, CA 94025 (US). CORLEY, Neil, C. [US/US]; 1240 Dale Avenue, #30, Mountain View, CA 94040 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). YANG, Junming [CN/US]; 7136 Clarendon Street, San Jose, CA 95129 (US). SHIH, Leo, L. [US/US]; Apartment B, 1081 Tanland Drive, Palo Alto, CA 94304 (US).	
(30) Priority Data:		(74) Agents: BILLINGS, Lucy, J. et al.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).	
60/172,216 20 October 1998 (20.10.98) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
60/118,559 4 February 1999 (04.02.99) US		Published	
60/172,229 11 February 1999 (11.02.99) US		With international search report.	
60/154,336 22 April 1999 (22.04.99) US		(88) Date of publication of the international search report: 28 September 2000 (28.09.00)	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications			
US 60/172,216 (CIP)			
Filed on 20 October 1998 (20.10.98)			
US 60/118,559 (CIP)			
Filed on 4 February 1999 (04.02.99)			
US 60/172,229 (CIP)			
Filed on 11 February 1999 (11.02.99)			
US 60/154,336 (CIP)			
Filed on 22 April 1999 (22.04.99)			
(71) Applicant (for all designated States except US): INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).			
(54) Title: PROLIFERATION AND APOPTOSIS RELATED PROTEINS			
<pre> 1  MSRTMARTRPGLG--RVTGAGGWGSAAVC 1342011 1  MATPVPPPSLRHLRLRLRLLSG-----LI GI452276  29  RGRALRGREPALPSASFPDVAACPGLDCA 1342011 25  LGALALNG-----ATARRPDAATTCPGSLDCA GI452276  59  LKRRARCFFGAHACGPGCLQPFQEDQOGLCV 1342011 50  LKRRARKCFFGAHACGPGCLQSFQEDQRFICY GI452276  89  PRMRFPFGGGRFPQPRLEDEIDFLAQELA-- 1342011 80  PRKHLSSGELPQPRLEDEIDSLAQELALK GI452276  117  RKEGHS--TPPLPKDRQRLPEPA-TLGF 1342011 110  EKEAGHSRLTAQPLLERAQKLLEPAATLGF GI452276  143  SARGQGLELGLPSTPGTPTPTPHTSLGSPV 1342011 140  SQWGORLEFLGLPSTHGTSLEIPTHSLSSRA GI452276  173  SSDPVHMSPLPFRGGQGDGLALVLI LAFVCV 1342011 170  SSGLPVQMSPLPFRGGQGDGLALVLI LAFCL GI452276 </pre>			
(57) Abstract			
<p>The invention provides human proliferation and apoptosis related proteins (PROAP) and polynucleotides which identify and encode PROAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of PROAP.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## INTERNATIONAL SEARCH REPORT

International Application No

PCT, US 99/24511

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C12Q1/68 A61K38/17 C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A61K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GALIANA E ET AL: "Identification of a neural-specific cDNA, NPDC-1, able to down-regulate cell proliferation and to suppress transformation." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, (1995 FEB 28) 92 (5) 1560-4., XP002133326 figure 2	5
A	--- DESBARATS, L. ET AL.: "Myc: a single gene controls both proliferation and apoptosis in mammalian cells" EXPERIENTIA, vol. 15, no. 52, December 1996 (1996-12), pages 1123-1129, XP000882427 the whole document --- -/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

29 March 2000

Date of mailing of the international search report

03. 07. 00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Herrmann, K

## INTERNATIONAL SEARCH REPORT

International Application No

PCT, US 99/24511

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 45128 A (APOPTOSIS TECHNOLOGY INC) 4 December 1997 (1997-12-04) the whole document ---	
A	WO 98 11256 A (LEVINE BETH C ;UNIV COLUMBIA (US)) 19 March 1998 (1998-03-19) the whole document ---	
A	WO 98 05347 A (YANAGISAWA JUNN ;SATO TAKA AKI (US); UNIV COLUMBIA (US)) 12 February 1998 (1998-02-12) the whole document -----	



# INTERNATIONAL SEARCH REPORT

Int. application No.  
PCT/US 99/24511

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claims 19 and 20 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 17, 18, 20 (completely), 1-16, 19 (all partially)  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
  
1 - 20 all partially

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17, 18, 20 (completely), 1-16, 19 (all partially)

Claims 17, 18, 20:

Compounds as such are not sufficiently defined by their mode of action. Therefore, claims 17, 18 and consequently claim 20 have not been searched because agonists and antagonist are neither disclosed nor supported within the terms of Art. 5 and 6 PCT, respectively. Example XIV on p. 57 of present description is not suitable to render said claims allowable under Art. 5 or 6 PCT.

Claims 1-20:

The fragments of claim 1 (polypeptide) and claim 9 (polynucleotide), respectively, are not defined by any clear technical (structural or functional) feature (Art. 6 PCT, Art. 84 EPC). A fragment can be as small as one amino acid or one nucleotide, respectively. Due to the infinite number of possible fragments a full search could not be carried out over the whole of the claimed scope. Therefore, the search for claims 1, 9 and 12 and all depending claims has been restricted to the complete sequences defined in the respective SEQ ID Nos.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: Invention 1: claims 1-20 all partially

Polypeptide comprising an amino acid sequence as in SEQ ID NO:1 and subject-matter relating to SEQ ID NO:1.  
Polynucleotides encoding the polypeptide of SEQ ID NO:1 such as a polynucleotide comprising a polynucleotide sequence as in SEQ ID NO:20 and subject-matter relating thereto.

2. Claims: Inventions 2-19: claims 1-20 all partially

Idem as subject 1 but limited to each of the polypeptides as in SEQ ID NOs:2-19 and polynucleotides as in SEQ ID NOs:21-38, respectively.

Invention 2 is limited to subject-matter relating to SEQ ID NOs 2 and 21 (amino acid and nucleic acid sequence of PROAP-2), invention 3 to SEQ ID NOs 3 and 22, etc.

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 99/24511

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9745128 A	04-12-1997	US 5834234 A	10-11-1998
		EP 0910387 A	28-04-1999
		US 6057132 A	02-05-2000
WO 9811256 A	19-03-1998	US 5858669 A	12-01-1999
		AU 4976997 A	02-04-1998
WO 9805347 A	12-02-1998	AU 4042497 A	25-02-1998
		CA 2260742 A	12-02-1998
		CN 1230120 A	29-09-1999
		EP 0935467 A	18-08-1999